DOCUMENT RESUME

ED 260 455 CS 209 283

AUTHOR Scherer, Darlene Lienau
TITLE Measuring the Measurements: A Study of Evaluation of Writing: An Annotated Bibliography.
PUB DATE 1 Aug 85
NOTE 77p.; Exit project, Indiana University.
PUB TYPE Reference Materials - Bibliographies (131) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Educational Assessment; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; *Holistic Evaluation; *Measurement Techniques; *Writing Evaluation; *Writing Research; Writing Skills

ABSTRACT

Intended to make the educational community aware of how research has defined acceptable practice in writing assessment, this annotated bibliography examines research about writing evaluation. Divided into five sections, the first section of the bibliography surveys some psychological and linguistic studies of the development of students' writing skills. The second section presents studies that describe methods of writing evaluation used in large-scale writing assessments. The focus of the third section is on research analyzing the writing task given to students in writing tests, with specific attention to the way the task was worded, the amount of explanatory information included in the task, the audience to which the writing was addressed, the mode of writing the student was asked to produce, and the method of presentation of the writing job to the students. The fourth section presents studies analyzing factors that produce bias in evaluators of writing, including characteristics found in the writing samples as well as characteristics found in the writers. Finally, the fifth section surveys research about grading methodology or response criteria, beginning with an analysis of experts' suggestions for organizing the grading process, and proceeding to studies comparing the direct and indirect methods of evaluation. The bibliography concludes with a summary of the research. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED260455

MEASURING THE MEASUREMENTS:

A STUDY OF EVALUATION OF WRITING

AN ANNOTATED BIBLIOGRAPHY

BY

DARLENE LIENAU SCHERER

Graduate Student

Study Directed

By

Charles R. DuVall

Submitted in partial fulfillment of the requirements
of the course E591---exit requirement
Indiana University at South Bend
August 1, 1985

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Darlene L. Scherer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

CS209283

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Statement of the Problem	3
Purpose of the Study	3
Plan of This Investigation	3
Definition of Terms	6
ANNOTATIONS	17
I. Development of Writing Skills	17
A. Psychological Studies	17
B. Linguistic Studies	19
II. Descriptions of Methods of Writing Evaluation	23
A. Indirect Assessment	23
B. Direct Assessment	24
1. Holistic Scoring	24
2. Primary Trait Scoring	28
3. Analytic Scoring	29
4. Atomistic Scoring	30
III. Research Related to Construction of the Writing Task	32
A. Procedures for Designing Prompts	32
B. Inclusion of Content in Prompts	33
C. Prompt Specification of Mode and Audience	34
D. Presentation and Interpretation of Prompt	36
IV. Factors Influencing Evaluators of Writing Samples	38
A. Evidence of Difference in Rating	38

Page

B. Factors Relating to the Writing Sample..... 38

C. Factors Relating to Teacher Expectations of the Writer..... 40

V. Research Relating to Response Criteria for Writing Evaluation..... 41

A. Method of Administering Response Criteria..... 41

B. Validity of Atomistic Response Criteria..... 43

C. Comparisons of Other Direct and Indirect Response Criteria..... 44

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS..... 47

Summary of Research..... 47

I. Development of Writing Skills..... 47

II. Descriptions of Methods of Writing Evaluation..... 48

III. Research Related to Construction and Use of Writing Prompt..... 51

IV. Factors Influencing Evaluators of Writing Samples..... 53

V. Research Relating to Response Criteria for Writing Evaluation..... 54

Conclusions..... 57

Recommendations..... 61

BIBLIOGRAPHY..... 63

INTRODUCTION

In 1975, the National Association for Educational Progress (NAEP) shocked the nation with the results of its studies: American high school students could not write well. Many more students wrote essays poorly in 1974 than did so in 1969. The December 1975 cover of Newsweek claimed to have an explanation of "Why Johnny Can't Write" (Fadiman, 1979, pp. 10-16, p. 146). Communities demanded that student writing skills be publicly demonstrated. In an effort to determine the writing skill level of their students, many localities instituted special tests, called competency tests. Students who passed the tests were declared competent writers; students who did not were to be given remediation and threatened with the ultimate weapon, denial of the high school diploma. Passing a writing proficiency test became a requirement for high school graduation in California, Georgia, New York, and Texas. The College Entrance Examination Board (CEEB) reinstated a written essay in the English Composition Test, many colleges instituted a written essay to determine placement in English classes, and some colleges and universities administered writing proficiency tests to would-be graduates (Freedman and Calfee, 1983, p. 75).

At the same time that the public demand for writing tests became insistent, educators realized that the measurement instruments for assessing writing competency were not

sufficiently developed to do the job. Jane Flinn (1984) remarked that "the nationwide cry for accountability in education has led to a distressing array of bureaucratic programs, poorly-designed tests, and punitive evaluation measures" (p. 101). UCLA writing researcher Edys Quellmalz (1982) diagnosed the problem accurately:

> Mounting concern for student achievement in writing has refocused attention on the features of writing assessment necessary to represent a student's skill fairly, usefully and economically . . . . Current practices are increasingly criticized regarding relevance to realistic writing situations, utility for forming decisions about individual competence or program effectiveness, fairness and legality for sanctioning exit requirements. (p. 1).

The scholarly educational community was forced to take a long, hard look at instruments for evaluating writing skill.

This long, hard look resulted in many research studies which focused on writing evaluation during the late 1970's and early 1980's. Two leading centers for this research were at the University of California at Berkeley, headquarters for the Bay Area Writing Project (BAWP) and the Center for the Study of Evaluation (CSE) at UCLA. A project currently in progress is the Ypsilanti Writing Research Project (YWRP), which has collected writing samples from elementary school children and is now engaged in psycholinguistic analysis of the data for future publication. (Tamor and Bond, 1983).

This paper will examine the research about writing evaluation produced at these centers and elsewhere in the

nation. Evidence found in these studies will enable educators to construct competency measurements which will evaluate student writing fairly and accurately.

STATEMENT OF THE PROBLEM

Current research is attempting to describe measurement practices which will allow writing achievements of children and youth to be evaluated fairly, reliably, validly, and economically. This study will evaluate the findings of this research in an effort to delineate exemplary evaluation practices.

PURPOSE OF THE STUDY

The purpose of the study is to make the educational community aware of how research has defined acceptable practice in writing evaluation.

PLAN OF THIS INVESTIGATION

Part I of this study will survey some psychological and linguistic studies of the development of students' writing skill. The psychological literature describes children's behavior in learning to write and the linguistic literature describes writing samples produced by young people. These psychological and linguistic developmental studies of writing acquisition must be included in a study of writing evaluation, because writing is a skill which is

acquired very slowly over a period of twenty or more years. Thus, if educators are going to get a valid measurement of this skill, it is important that an informed picture of their subjects' developmental stages be a part of the formulation of the evaluation task and the response to it.

Part II of the study will abstract literature which describes methods of writing evaluation used in large-scale writing assessments. The methods of writing evaluation examined are first, indirect methods, in which the student selects answers on an objective or multiple choice test and second, direct methods, in which the student produces a sample of writing which is evaluated by raters. Direct measures which will be described are: holistic scoring, primary trait scoring, analytic scales, and atomistic ratings.

Part III of the paper will focus on research analyzing the assignment given to students in a writing test, called the writing task. These studies investigated the way the prompt, or the writing assignment, was worded, the amount of explanatory information included in the prompt, the audience to which the writing is addressed, the mode or genre of writing the student was asked to produce, and the method of presentation of the writing job to the students.

Part IV will present studies which analyze factors producing bias in evaluators of writing. These factors included characteristics found in the writing samples as well as characteristics found in the writers.

Part V of the study surveys research about grading methodology, called in the literature, response criteria. It begins with an analysis of experts' suggestions for organizing the grading process. It then proceeds to studies comparing the direct and indirect methods of evaluation in terms of reliability, validity, correlation to other methods, and economy.

DEFINITION OF TERMS

Analogic Writing -- a term used by James Britton to signify a type of informative writing in which generalizations are related hierarchically or logically (Britton, 1975, backpiece).

Analogic Writing: Low Level of Generalization -- a term used by James Britton to signify a type of informative writing which contains generalizations which are loosely related (Britton, 1975, backpiece).

Analogic-Tautologic Writing -- a term used by James Britton to signify a type of informative writing which uses speculation about generalizations; the open-ended consideration of analogic possibilities (Britton, 1975, backpiece).

Analytic Scale -- a method of evaluation of writing samples in which prominent features of the writing are scored separately and the final mark is a tally of the individual scores (Cooper, 1977, p. 7).

Anchor Paper -- a paper selected from within the writing samples produced in an assessment, which defines a scoring category (Myers, 1980, p. 33).

Atomistic Writing Test -- a test of writing ability which relies on the assessment of particular features associated with skill in discoursing. These tests quantify the smallest units of discourse found in a piece of writing, such as vocabulary, usage, manuscript mechanics, syntactical or rhetorical

units for measuring maturity or clarity in discourse (Lloyd Jones, 1977, p. 33).

Audience -- those who pay attention to what one writes or says; in writing, the person(s) to whom the writing is addressed, the intended reader (Webster, 1980).

BAWP -- acronym for the Bay Area Writing Project which centered around the Graduate School of Education, University of California, Berkeley (Myers, 1980, p. 2).

Bias -- the influence on grades of irrelevant considerations such as liking or disliking the student, disagreement with his views (Diederich, 1974, p. 99).

Cluster -- a group of readers whose grades agree within their group and disagree with the grades of every other group to a greater extent than can be attributed to chance (Diederich, 1974, p. 99).

Co-Referencing -- a linguistic device for constructing coherent text. Use of pronouns and lexical repetitions establish co-referencing. Sometimes there is a need to add new differentiating information to the text (Bartlett, 1981, p. 20).

Competency Testing -- a procedure for determining what a student knows or can do relative to some standard of performance. Competency tests are designed to identify students who fail to meet a certain

standard (Cooper, 1981, p. 2).

Construct Validity -- the extent to which test performance can be interpreted in terms of certain explanatory concepts which account for the evaluation of a performance. In writing, the subscales on an analytic scale might be the factors in the construct (Dilworth, 1979, pp. 43-44).

Content Validity -- refers to the extent to which a test measures a representative sample of the domain of tasks under consideration (Gronlund, 1981, p. 90).

Correlation -- a mathematical procedure that shows to what extent it is true that, the higher a student stands on one measure, the higher he stands on another. The measures need not be of the same characteristic nor on the same scale, but one must correlate two sets of measures of the same students (Diederich, 1974, p. 99).

Criterion Validity -- the extent to which a score relates to another variable that is a direct manifestation of the characteristic in question. A composition grade must be translatable to specific behaviors that characterize writing performance at the level of that grade (Dilworth, 1979, p. 44).

CSE -- an acronym for the Center for the Study of Evaluation headquartered at the Graduate School of Education, UCLA (Burry and Quellmalz, 1983, p. 1).

Direct Assessment of Writing -- evaluation in which the

student produces a writing sample which is then assessed for qualitative features (Hall, 1981, pp. 2-5).

Discourse -- a communication of thoughts by words; expression of ideas in writing (Webster, 1980).

Discrepant Grades, Review of -- a procedure in which essays that received far different grades from the original readers were referred to a more experienced reader for re-evaluation. The third grade was substituted for the grade which was farthest from it (Diederich, 1974, p. 101).

Essay Scale -- a type of holistic evaluation of writing. The essay scale is a series of complete pieces arranged according to quality. At one end of the series is an exemplary piece, at the other an inadequate one. The pieces which make up the scale are usually selected from pieces written by students like those with whom the scale will be used. A rater attempts to place a new piece of writing along the scale, matching it with the scale piece most like it (Cooper, 1977, p. 4).

Evaluation -- the term may be used in two ways: 1) measurement, or the quantitative description of pupil behavior, such as test scores, or 2) those appraisal methods that do not depend on measurement—the qualitative descriptions of pupil behavior. From an instructional standpoint evaluation may be

defined as the systematic process of determining the extent to which instructional objectives are achieved by pupils. Evaluation always includes value judgments concerning the desirability of the results (Gronlund, 1981, pp. 5-6).

Exposition -- writing that intends to inform by presenting facts, giving directions, recording events, interpreting facts, or developing opinions (Burry and Quellmalz, 1983, p. 8).

Expressive Language -- writing close to the self, revealing the speaker verbalizing his/her consciousness, displaying a close relationship with the reader. It is relatively unstructured and may possibly not be highly explicit. This is a term used by Britton as a function category of language use (Britton, 1975, backpiece).

General Impression Marking -- a method of holistic evaluation in which the rater quickly reads the paper and scores it, based on his overall impression of the quality of the work (Cooper, 1977, pp. 11-12).

Generalized Narrative or Descriptive Information -- a term used by James Britton to signify a type of informative writing in which the writer is tied to particular events and places but is detecting a pattern of repetition in them. He expresses this in a generalized form (Britton, 1975, backpiece).

Genre -- a term used in literary or rhetorical criticism

to designate the distinct types or categories into which discourse is grouped according to form or technique. Genre classification implies that there are groups of formal or technical characteristics existing among works of the same kind (Thrall, Hibbard, and Holman, 1960, p. 211).

Grade -- to assign a conventional and public mark to evaluation (Cooper, 1975, p. 14).

Holistic Evaluation of Writing -- rating on general impression. It consists of giving a single grade or score to an essay rather than a number of ratings on various qualities (Diederich, 1974, p. 100).

Indirect Assessment of Writing -- students are evaluated by taking an objective test which measures discrete subcomponents of writing, most commonly vocabulary, grammatical usage, capitalization, punctuation, spelling, paragraph organization, and sentence structure (Hall, 1981, pp. 2-5).

Mechanics -- the mechanical aspect or technical part of writing (Webster, 1980). In the Diederich Expository Scale, mechanics included word usage, punctuation, spelling and handwriting (Diederich, 1974, p. 54).

Mode -- the method or form of the expression of an idea (Webster, 1980). Some writing modes are exposition, persuasion, narration, or description.

NAEP -- an acronym for National Assessments of Educational

Progress, which were conducted in 1969, 1974, and 1979. It is a federally funded program conducted by the Education Commission of the States (Fadiman, 1979, p. 13).

Narration -- a description of an event(s), based on an account. It includes such aspects as 1) how, when, and where the events took place; 2) the persons, places, or things involved in the events; and 3) the actions, thoughts, or feelings of the actors described (Burry and Quellmalz, 1983, p. 10).

Participant Use of Language -- a use of language in either writing or speech, which is characterized by the need to act and decide. Some participant uses of language are informing, instructing, persuading, arguing, explaining, planning, coming to a conclusion. Language used in the participant role is selected and ordered according to demands made outside the writer (Britton, 1970, p. 122-4).

Poetic Language -- a verbal construct; patterned verbalization of the writer's feelings and ideas. This category is not restricted to poems but would include such writings as the short story, play, autobiographical episode. This is a term used by James Britton as a function category of language use (Britton, 1970, backpiece).

Primary Trait Scoring -- a method of writing evaluation in which the rater focuses on just those features of

a piece of discourse which are relevant to the kind of discourse it is. The scoring guides are constructed for a particular writing task set in a full rhetorical context (Cooper, 1977, p. 11).

Prompt -- the invitation to write in a writing test, the test question, the writing stimulus. The writing assignment usually consists of: 1) the topic or subject, 2) the instructions or suggestions about or limitations on the content of the writing and how the writer is to address the subject (Gray, 1982, p. 135).

Referential Ambiguity -- a kind of non-standard syntax in written discourse in which the references for pronouns are not expressed in the preceding text (Bartlett, 1981, p. 19).

Reliability -- the consistency of evaluation results. If a test is reliable, students will obtain similar scores when the same test is administered to the same group on two different occasions. If different teachers independently rate the same pupils on the same instrument and obtain similar ratings, we can conclude that the results have a high degree of reliability (Gronlund, 1981, pp. 65-66).

Response Criteria -- the manner in which a writing sample is graded or rated, the particular method used to evaluate a written script (Winters, p. 3, 1978).

Rubric -- printed directions to writing evaluators describing

the use of a particular method of scoring (Myers, 1980, p. 51).

Spectator Use of Language -- a use of language in which the writer or speaker describes experience, ideas, or feelings with a freedom to evaluate more broadly, relating the event to a system of values, knowledge, or beliefs, and savors feeling as feeling. In the written language in the role of spectator the arrangement of forms and the response of a reader to the arrangement of forms are essential features. Uses of language in the spectator role include story-telling, fiction, the novel, drama, and poetry (Britton, 1975, pp. 109-115).

Staff Grading -- the grading of test essays by all members of the department after some training in grading by common standards. The essays are identified only by code numbers so that no reader knows which student wrote any paper. Each essay is graded by two readers, who record their grades and comments on separate work sheets and write nothing on the essays themselves, so that no reader knows which grade any essay received from any other reader (Diederich, 1974, p. 101).

Syntactic Fluency -- the ease and skill with which a writer uses sentence-combining transformations to include within each full sentence all ideas one wishes to include, cast in forms that are stylistically

appropriate and grammatically correct. The degree of syntactic fluency depends upon short-term memory and the willingness to reflect upon and craft utterances (Cooper, 1981, p. 40).

T-Unit -- a single main clause (or independent clause) plus whatever other subordinate clauses or non-clauses are attached to, or embedded within, that one main clause (Hunt, 1977, pp. 92-93).

Tautologic Writing -- a term used by James Britton to signify a type of informative writing which uses hypotheses and deductions from them. In this form of writing theory is backed by logical argumentation (Britton, 1975, backpiece).

Transactional Writing -- language used to get things done. It is concerned with an end outside itself. It informs, persuades and instructs. Term used by James Britton as a function category of language use (Britton, 1975, backpiece).

Validity -- the extent to which the results of an evaluation procedure serve the particular uses for which they are intended (Gronlund, 1981, pp. 67-69).

Writing -- composing a piece of discourse which involves thinking and languaging. Composing a piece of writing is an experience itself. The act of writing-down contributes to the actual shape and meaning of the finished piece. Writing is a way of ordering and clarifying experience. It occurs in different

forms for different purposes or audiences (Cooper, 1975, p. 113).

UNIVERSITY OF NOTRE DAME

ANNOTATIONS

I. DEVELOPMENT OF WRITING SKILLS

A. Psychological Studies

Fadiman, Clifton, and James Howard. Empty Pages: A Search for Writing Competence in School and Society. Belmont, Calif.: Fearon Pitman, 1979.

Children first learn language from speech they hear in their own homes. Conversation with parents and listening to parents read aloud helps children learn to put words together and instills a feeling for the way words work. A condition of learning to write is incentive. The young person must know or feel that s/he can accomplish some important need by writing. Learning to write takes a long time. A writer gains proficiency only as s/he assimilates what s/he hears, reads, observes, thinks, and does from day to day. Freedom to make mistakes is a most important condition of learning to write. A child may not be ready for editing until the fourth or fifth grade. If school gives young people plenty of opportunity to write, they are likely to make more and better progress than if writing is only an occasional activity. Teens, being somehow more directly affected by the so-called adult culture than younger children, are peculiarly susceptible to the conditions in society that discourage people from writing. These conditions are mute families, immediacy-oriented media, and under-educated adult models.

Macintosh, Helen, and Wilhelmina Hill. How Children Learn to Write. Bulletin No. 2--Place of--Subjects Series. Washington, D.C.: U.S. Dept. of Health, Education, and Welfare, Office of Education, 1953.

The child's need for both purposeful and creative writing has a close relationship to his total growth pattern. When children write for a purpose, the idea or information to be communicated is of primary importance. Children need guidance in the selection and clarification of the ideas to be written. When teachers and parents see a child's writing achievements in terms of his own interests and needs and his total growth pattern, they are able to give him more effective assistance in making appropriate progress. Children who encounter adult disapproval and penalties often develop blockages to learning the skills of written expression, which then are greatly delayed or never attained. Practice in mechanical writing skills should follow rather than precede use in situations that are meaningful. Children learn to write by writing. There should be opportunities for every child to write every day in ways that are purposeful.

Britton, James. "Now That You Go to School." In Language and Learning. London: Allen Lane The Penguin Press, 1970.

The beginning of school leads to the development of differences in language usage, a continuation and refinement of language in the role of participant. The earliest forms of written down speech are likely to be expressive. The demand for participant language that any reader can follow or for spectator role language to satisfy an unknown reader pressures the child to move from expressive writing to transactional and poetic writing. Stories and poems children hear and read in school extend the world representation of the child. It is important that the child unites language heard in reading to his spoken language in the process of composing written language. When children use language in school, it must be for operations and not dummy runs.

Britton, James. "The Question of a Lifetime." In Language and Learning. London: Allen Lane The Penguin Press, 1970.

The adolescent task of becoming an integrated individual and establishing links between his many social roles is reflected in adolescent writing. Most of this writing is in the spectator role in such forms as autobiographical accounts, stories, and poems. A large part of the incentive for the writer lies in the sharing of experiences with peers. Through their writing adolescents begin to achieve a more mature view of themselves and the world by looking back into their own childhoods. It is the achievement of late adolescence to theorize objectively, to handle highly abstract concepts with due regard for their logical relationships at a concrete or empirical level. Writing in the participant role, the developing young person will argue, confess, explore, theorize. With the achievement of mature powers of thinking, the adolescent is able to construe things as they might be. Adolescents revolt against the language of authority figures around them because these formulations begin to conflict with their own responses to changing situations. Therefore, development of the appropriate forms of speech is likely to come fairly late in adolescence.

B. Linguistic Studies

Britton, James, Tony Burgess, Nancy Martin, Alex McLeod, and Harold Rosen. "The Process of Writing." In The Development of Writing Abilities (11-18). London: Schools Council Publications, Macmillan Education Ltd., 1975.

The beginning step in the writing process is conception, in which the writer deliberately selects from her previous experience and relates her task to her own hierarchical construct system, gleaned from past experience of writing in school and the way the teacher sets up the writing task. The next step, incubation, includes writer understanding of the idea and a plan for its expression. Incubation can take much time, yet children are constantly judged on impromptu and instantaneous performances. The final step in the writing process, production, is crucial psychologically. Getting started writing involves finding the right tone to talk with and finding a way into the topic. While writing, scanning back is needed to help the writer keep overall control of what she is doing. Fluent writers draw on integrated movement sequences to write large numbers of words as meaning units. One constraint on the writing process is the number of words that can be held in the short-term memory at one time. When one writes, the short-term memory draws on the long-term memory as successive words appear on the page. For a child just learning to write, the number of words in the short-term memory may be small.

Bartlett, Elsa Jaffe. "The Second 'R'." In Learning to Write: Some Cognitive and Linguistic Components. Papers in Applied Linguistics, Linguistics and Literacy Series: 2. Washington, D.C.: Center for Applied Linguistics, 1981.

Writing begins with an intent to accomplish something and these intentions are of two sorts: 1) to have some intended effect on readers and, 2) to produce a certain type of artifact. At the same time, wordings must be chosen that will enable readers to link the incoming information to what they already know about the text. The writer must also manage information so that a reader has sufficient context to interpret a particular piece of text. Writing is both an ad hoc and a patterned kind of activity. Therefore, two curricular strategies will prove useful. First, beginners will become better at solving composing problems if their assignments and discussions are organized around the structural properties of text. Second, young writers will benefit from activities designed to make these properties available for conscious reflection and articulation.

Britton, James, Tony Burgess, Nancy Martin, Alex McLeod, and Harold Rosen. "Results: The Writer's Sense of Audience." The Development of Writing Abilities (11-18). London: Schools Council Publications, Macmillan Education Ltd., 1975.

This study, which began in 1965, analyzed students' sense of audience evident in their writing. Audience classification categories included: 1) self, 2) teacher, 3) known wider audience, 4) unknown audience. Teacher-raters analyzed 2,122 pieces of written work from several subject areas written by 500 boys and girls ages 11-18 in various English secondary schools. The overall tendencies of the writing sample suggested the dominance of the teacher audiences (85%), and in particular those of pupil-to-examiner (48%) and teacher-learner dialogue (38.8%). In the first and third years the writing was totally dominated by teacher audiences. The proportion of writing for the teacher-as-examiner increased somewhat in the third year, while writing for the teacher-learner dialogue correspondingly declined. A more marked change came with the fifth year, when writing both for an examiner audience and for audiences other than the teacher increased, while writing for the teacher-learner dialogue decreased sharply. The pattern was continued through into the seventh year.

Britton, James, Tony Burgess, Nancy Martin, Alex McLeod, and Harold Rosen. "Results: Function." In The Development of Writing Abilities (11-18). London: Schools Council Publications, Macmillan Education Ltd., 1975.

This study, which began in 1965, undertook a developmental analysis of the processes by which the written language of children aged 11-18 became differentiated into kinds of written discourse appropriate to different purposes. The discourse function categories classification included: 1) Transactional (Informative or Cognative), 2) Expressive, 3) Poetic, and 4) Immature Categories. The overall tendencies of the sample suggested the dominance of transactional writing, and within that category, of the analogic level of the informative subdivision. Analysis of the subtotals in the four age groups represented confirmed this pattern, revealing a developing increase in analogic writing which was substantially accelerated in the fifth and seventh years. While there was evidence of development in the ability to use writing with higher levels of abstraction, the incidence of speculative or tautologic writing was comparatively low. The evidence of writing in functions other than the transactional was slight. Expressive and poetic writing were minimally represented throughout the sample.

Bartlett, Elsa Jaffe. "Text Concept and Discourse Plans." In Learning to Write: Some Cognitive and Linguistic Components. Papers in Applied Linguistics, Linguistics and Literacy Series: 2 Washington, D.C.: Center for Applied Linguistics, 1981.

Research on narratives suggested that while some initial learning seems to occur fairly early, knowledge of these forms continues to develop rather slowly throughout the elementary and junior high years. Data on expository writing suggested an even lengthier period of basic development that extends into high school and college. Development of skill in writing any particular type of text consists, in part, of an increasing ability to represent that text in terms of its structural features and to compose in terms of those features. Several types of experiences would be helpful in fostering the development of these skills and concepts. Students should read in the various text forms. Students should then consider these texts in terms of their structural features. Awareness of form may be heightened if beginning writers are given opportunities to try their hands at composing some of the more highly patterned narrative and expository formats.

Bartlett, Elsa Jaffe. "Coherence and Wording." In Learning to Write: Some Cognitive and Linguistic Components. Papers in Applied Linguistics, Linguistics and Literacy Series: 2. Washington, D.C.: Center for Applied Linguistics, 1981.

Research data suggested that elementary, junior high students, and possibly some adults can revise a tester-prepared text, but have difficulty revising their own text, due to the inability to put aside intended meaning and interpret text with just the information actually provided. In one experiment, fifth graders could detect referential ambiguities in the texts of others, but not in their own. In other experiments on fourth, sixth, and seventh graders, students were unable to fix up referential ambiguities when differentiating information was not provided in the context. Another study found that third, sixth and ninth graders had difficulty incorporating new information into a text, but instead added it on at the end of the text.

Odell, Lee. "Redefining Mature Writing." Albany: State Univ. of New York, 1979. Paper read at the Annual Meeting of the Canadian Council of Teachers of English, 8-11 May, 1979, Ottawa, Canada. (ERIC Document Reproduction Service No. ED 173 781).

Odell analyzed a variety of mature and immature writing samples produced by youthful and grown-up writers alike, to deduce the following characteristics of truly mature writing: 1) it recognized that the audience is different from the writer; 2) it provided an appropriate context for his or her statements; 3) it tried to anticipate and respond to objections or questions the audience was likely to have; 4) it recognized the legitimacy as well as the limitations of other points of view on a given subject; 5) it acknowledged, where appropriate, the limitations of the writer's own point of view; 6) it indicated what the viewpoint could not explain; 7) it took note of apparently contradictory evidence; and 8) it recognized the complexity of the subject at hand by attending to more than one feature of an experience.

Hunt, Kellogg W. "Early Blooming and Late Blooming Syntactic Structures." In Evaluating Writing: Describing, Measuring, Judging. Ed. Charles R. Cooper and Lee Odell. Urbana, Ill.: National Council of Teachers of English, 1977.

This study described syntactic structures which appear in developmental sequence. T-unit coordination appeared early, but gradually was replaced by more sophisticated structures. Coordination between predicates appeared at grade four but remained constant thereafter. As school children got older, their t-units lengthened and incorporated more complex structures. Students in grade six and older were able to coordinate objects. Use of the appositive appeared in grade eight. The ability to take an adjective out of the nominative position and use it as an adjective began in sixth grade and increased steadily to grade twelve. Older students made syntactic category shifts more readily than younger. Using verbs as participial phrases showed up on the college level. These syntactic maturity studies were replicated in the Netherlands, Fiji, Korea, Indonesia, Laos, and Japan.

Benton, S.L., Robert G. Kraft, John A. Glover, and Barbara S. Plake. "Cognitive Capacity Differences Among Writers." Journal of Educational Psychology 76 (1984): 820-34.

Differences between good and poor writers were examined on six cognitive-processing tasks. In Experiment 1, an ordered letters task, an iconic memory task, and a letter reordering task were employed to ascertain differences between good and poor undergraduate writers. In Experiment 2, three additional tasks were employed to examine good and poor writers' abilities to hold and manipulate larger amounts of information: a word reordering task, a sentence reordering task, and a paragraph assembly task. Experiment 3 replicated the results of Experiments 1 and 2 with a high school sample. The results of the studies suggested that good writers manipulated information in short term memory significantly more rapidly than poor writers. Further, this greater rapidity of information manipulation did not seem to occur at the expense of accuracy. It was suggested that the observed differences between good and poor writers could be attributed to differences in elementary information-processing systems. Good writers seemed to have developed whole series of programs that operated almost automatically as writing occurred.

II. DESCRIPTIONS OF METHODS OF WRITING EVALUATION

A. Indirect Assessment

Hall, Janice K. Evaluating and Improving Written Expression: A Practical Guide for Teachers. Boston: Allyn and Bacon, 1981.

Standardized achievement tests measure discrete subcomponents of writing, most commonly vocabulary knowledge, grammatical usage, capitalization, punctuation, and spelling. Occasionally, achievement tests will include items on topic sentences, sequencing sentences, and eliminating irrelevant sentences. These skill areas are tested by proofreading and editing tasks in multiple-choice format. Students are asked to pick the correct word to complete the meaning of a sentence, to choose the best sentence from a group of sample sentences, to supply the correct capital letter or punctuation symbol, or to choose the correct spelling of a word. Some objective measures utilize practical writing samples as stimuli for multiple-choice questions. These subcomponents of writing have been selected because they can be easily standardized and scaled in relation to representative populations. Test results are provided in the form of grade equivalents and percentiles. There is general agreement that standardized achievement tests do not provide enough information about students' actual writing ability. Achievement tests are considered inadequate measures of writing ability because students do not produce a complete writing sample.

Braddock, Richard. "Evaluation of Writing Tests." In Reviews of Selected Published Tests in English, Ed. Alfred H. Grommon. Urbana, Ill.: National Council of Teachers of English, 1976.

This article described the content of standardized tests as consisting almost entirely of items concerned with sentence structure, word meaning, and mechanics. Occasionally such a test will include items which relate to a student's ability to judge the organization and substantive details of a piece of writing. Standardized tests were criticized for not sampling a student's ability to write standard English, only to recognize it, and for not measuring a student's ability to select a subject and an approach and a mode for it appropriate to both writer and reader. These tests did not measure a student's ability to organize and detail his writing. The article evaluated four standardized writing tests which the author considered good examples of indirect measures: the College English Placement Test (CEPT); Content Evaluation Series Language Arts Tests, Composition Test; the McGraw-Hill Basic Skills System Writing Test; and the Cooperative Sequential Tests of Educational Progress (STEP).

B. Direct Assessment

1. Holistic Scoring

White, M. "Holisticism." College Composition and Communication 35 (Dec. 1984): 400-9.

Holistic scoring emphasizes evaluation and response to student writing as a unit without sub-scores or separable aspects. It is opposed to multiple-choice testing and the analytic approaches to writing. Holisticism in grading reflects the view that writing activities are not describable through an inventory of their parts. Holistic scoring is able to achieve acceptably high reliability by adding a series of constraints to the economically efficient practice of general impression scoring. These constraints, which attempt to reduce unnecessary variability in the scoring process, are: 1) a controlled essay reading, 2) a description of papers at different points on the scoring scale, 3) sample papers or anchors, 4) checks on the reading in progress by table leaders, 5) multiple independent scoring by different readers. Holistic scoring can not give diagnostic information beyond the comparative ranking it represents. The ranking is connected to a particular test group; it does not represent an absolute.

Freedman, Sarah Warshauer and Robert C. Calfee. "Holistic Assessment of Writing: Experimental Design and Cognitive Theory." In Research on Writing: Principles and Methods. Ed. Peter Mosenthal, Lynne Tamor, Sean A. Walmsley. New York: Longman, 1983.

Holistic scoring was defined as the evaluation process in which essays are given a numerical score which represents the rater's opinion of the overall quality of the essay. Raters were instructed to read the essay quickly to score it as a whole without pondering the parts. The first study determined the importance of three factors which influenced how raters graded essays holistically: the essay, the rater, and the context. The results showed that the essays themselves were the chief source of systematic variation, more than 80%. The findings suggested that skilled raters, when provided specific guidelines and supervised training, were able to rate a wide variety of compositions in a consistent fashion. The second study attempted to determine the extent to which each of the four categories of the essay—idea development, organization, sentence structure, and mechanics—caused raters to vary their holistic scores. The results showed that three of the four factors strongly influenced the raters' judgments; development had the greatest effect, followed by organization and mechanics.

Chew, Charles R. "A Writing Test: A Search for Competency: The New York Experience." Albany: New York State Dept. of Education, 1981. Paper read at International Conference on Language Problems and Public Policy, 16-19 Dec., 1981, Cancun, Mexico. (ERIC Document Reproduction Service No. ED 203 010).

This paper described the high school graduation writing competency requirements in New York state, which consisted of two tests: the Preliminary Competency Test in Writing (PCT), given to students in the eighth or ninth grade, and the Regents Competency Test in Writing (RCT) given to eleventh grade students. The RCT requires three writing samples—a business letter of complaint, a report from given data, and a piece of persuasion. The Dept. of Education provides schools in the state with a listing of 25 topics for each task. Evaluation of the writing samples is with exemplary model answers and a set of criteria derived from the exemplary models. Each of the three samples is rated by a different teacher, with the three ratings averaged to determine the test score. All tests with a 60% mean or above are rerated by the State Dept. of Education to maintain a uniform statewide graduation standard in writing.

Alloway, Evans. "Holistic Scoring Session for the New Jersey Writing Project: A Summary." Trenton, New Jersey, May, 1979. Report prepared by the New Jersey Writing Project, New Jersey State Dept. of Education. (ERIC Document Reproduction Service No. ED 191 069).

The holistic procedure used to score 2,800 student papers in the New Jersey Writing Project is described in the booklet. Characteristics were given of papers that received high, middle, and low scores. Papers which received high scores were characterized by a clear understanding of and a direct approach to the assignment, a well-presented opinion, a clear focus and appropriate organizational pattern, a sense of closure, a sense of style, and relative freedom from errors. Papers which received middle scores were characterized by inconsistency. Papers which received low scores contained a significantly greater number of simple sentences, lacked focus, and were often seriously flawed grammatically. A sample essay at each level was provided.

Spencer, Ernest. "Folio Assessment of External Examinations? An Investigation into Alternative Means of Assessing SCE Ordinary Grade English." Report to the Scottish Certificate of Education Examination Board from the Scottish Council for Research in Education, 1979. (ERIC Document Reproduction Service No. ED 194 555).

From 1974 through 1978, a study was conducted of three methods of assessing Scottish high school students' English achievement: 1) ordinary O-level examinations, 2) folio assessment of writing skills, 3) criterion referenced tests. For the folio assessment pupils' English composition work was retained in a folder and assessed by teachers of the school. Six of nine assignments in factual, persuasive, personal, descriptive, or poetic writing of 350-400 words or longer were included. The criterion test lasted 7½ hours and consisted of five essays—factual, persuasive, or literary interpretation included—plus a multiple choice test of reading comprehension. Results showed that the Folio Assessment, O-Grade Examination, and Criterion Test scores correlated with one another satisfactorily. General conclusions of the study favored the O-grade examination, because its nationwide use insured comparability of scoring standards. Problems with folio assessment were marked inconsistency and influence of class groupings on teacher marking.

Nyberg, Verner and Adell M. Nyberg. "Technical Report on Alberta Essay Scales; Models." Edmonton, Canada: University of Alberta, 1982. (ERIC Document Reproduction Service No. ED 220 484).

The models were chosen from essays submitted by 8,020 twelfth grade students in 1964. The essays were graded analytically at least eleven times on two different scales, one for mechanics and one for style content. Spelling, punctuation, word usage, grammar, sentence errors and form were considered in the mechanics scale. Organization, diction, sentence organization and total impression were considered in the style/content scale. Then model essays were selected for both mechanics and style content to represent stanines 2, 4, 6, 8. Essays judged poorer than the models were to be given the odd numbers. When using the models, raters were to match scripts to the model essays, using a holistic, impressionistic approach.

Bamberg, Betty. "Assessing Coherence: A Reanalysis of Essays Written for the National Assessment of Educational Progress, 1969-1979." Research in the Teaching of English 18 (Oct. 1984): 305-19. (ERIC No. EJ 305 318).

The purpose of this study was to develop a valid method of assessing essay coherence and to reanalyze essays written by 13-17 year olds for the 1969, 1974, and 1979 National Assessments of Educational Progress (NAEP). A Holistic Coherence Scale based on recent theory was developed and used to assess coherence of 2,698 essays. This scale assessed coherence holistically, in terms of a list of features that create both global and local coherence. It rated essays on a four point ordinal scale that conceptualized coherence as a quality achieved with varying degrees of success rather than as a dichotomous variable. The NAEP studies found a total decline in coherence between 1974 and 1979. However, this study found between-assessment difference was not as great as the difference found between 13 and 17-year-olds. Essays written by 17-year-olds were significantly more coherent than essays written by 13-year-olds in all three assessments, a result not found in NAEP's original analysis. Also, a strong correlation between essay coherence and quality was found.

2. Primary Trait Scoring

Lloyd-Jones, Richard. "Primary Trait Scoring." In Evaluating Writing: Describing Measuring Judging. Ed. Charles R. Cooper and Lee Odell. Urbana, Ill.: National Council of Teachers of English, 1977.

Primary Trait scoring defined writing accomplishment in terms of a particular segment of discourse. The steps in using the Primary Trait Scoring System were to define the universe of discourse, to devise exercises which sample that universe precisely, to ensure cooperation of the writers by pretesting the instrument, to devise a workable scoring guide and to use it. The scoring guide should include an interpretation of the exercise which indicates how each element of the stimulus is presumed to affect the respondent and a system for defining the shorthand which is to be used in reporting descriptions of the writing (most guides use numerical scores), and samples of papers which have been scored, including discussions of why each sample was scored as it was. Examples of narrative and persuasive assessments were given. The narratives were scored in terms of use of dialogue, point of view, and tense. The persuasive pieces were scored in terms of point of view, number and elaboration of reasons, and kinds of appeals.

Rabianski, Nancyanne. "Scoring Writing with an Informative Aim." Document reviewed in Measures for Research and Evaluation in the English Language Arts, Vol. 2, The Research Instruments Project. Urbana, Ill.: Committee on Research, National Council of Teachers of English, 1977. (ERIC Document Reproduction Service No. ED 236 638).

Based on the discourse theories of Kinneavy and Britton, this writing assessment method evaluated the degree to which an essay is successful in giving information. The judgment is determined by scoring each generalization with one of five numerical values according to the quantity of supportive elaboration employed, ranging from unelaborated through outstanding elaboration. The rating guides were based on an examination of 372 essays produced by 93 tenth graders. Inter-rater reliability for the total of all topics was .885.

3. Analytic Scoring

Diederich, Paul Bernard. "Descriptions of Papers Rated High, Middle, and Low on Eight Qualities." In Measuring Growth in English. Urbana, Ill.: National Council of Teachers of English, 1974.

The parts of the analytic scale were short names for separate judging techniques developed through factor analysis of grading done by many teachers and mature writers. Each aspect of the essay was scored separately, on a scale of 2-10 points for ideas and organization, and on a scale of 1-5 points for wording, flavor, usage, punctuation, spelling, and handwriting. The points were tallied for the final score of each paper. The scale also included a description of papers in the high, middle, and low level for each of the qualities rated. Sample typical papers which exemplified the high, middle and low ratings were chosen from the many papers rated to be used as guidelines for the teachers doing the rating.

Burry, James and Edys Quellmalz. "Assessing Students' Writing Skills: The CSE Expository and Narrative Rating Scales." Los Angeles: University of California, 1983. (ERIC Document Reproduction Service No. ED 238 942).

The CSE expository and narrative rating scales were developed to provide sound, instructionally relevant methods for assessing students' writing competence. Each scale's analytic rating elements were referenced to specific rhetorical features characteristic of the writing aim addressed. The elements in the narrative scale were first, general competence, which made an overall evaluation of the essay. The other subscales—focus/organization, support, and mechanics—made analytic evaluations of the specific components of narrative writing. The expository scale began with a rating of general competence in which raters judged whether the work achieved an expository purpose for the intended audience. The remaining subscales were analytic, including essay organization/coherence, paragraph organization, support, and grammar/mechanics. Each aspect of both scales was rated on a 1-6 point scale, with scores of 4-6 representing mastery and scores of 1-3 representing non-mastery.

Malgady, Robert G. and Peter R. Barcher. "Some Information Processing Models of Creative Writing." Journal of Educational Psychology 71.5 (Oct. 1979): 717-25.

Essays written by 38 eleventh graders were judged by 66 teachers in two ways: first on a holistic rating of creativity and then on a five-element creativity scale which included productivity, novelty, figures of speech, flexibility, and elaboration. The scores on individual aspects of the scale were compared with the holistic rating. Four of the five dimensions hypothesized to represent creativity corresponded significantly with teachers' global judgments of essay creativity. The only factor not significantly correlated with creativity was elaboration. There was 90% agreement between judges in scoring these dimensions.

4. Atomistic Scoring

Hunt, Kellogg W. Syntactic Maturity in School Children and Adults. Monographs of the Society for Research in Child Development, Serial No. 134, 35.1. Chicago: University of Chicago Press, 1970.

In 1965 the researcher analyzed the writings of children of average IQ in grades four, eight, and twelve, plus the writing of skilled adults, to develop the t-unit, a measure of syntactic maturity in writing. A t-unit, or minimal terminable unit, was defined as one main clause plus any subordinate clause or nonclausal structure that was attached to it or embedded in it. Any complex or simple sentence would be one t-unit, but any compound or compound-complex sentence would consist of two or more t-units. Hunt's analysis of these writings found: 1) as school children mature, they tend to write more clauses per t-unit; 2) words per t-unit increase with maturity; 3) as school children mature, they tend to write fewer t-units per sentence; 4) taking the maturity of the average twelfth grader as the goal, the best index of syntactic maturity is words per t-unit, second best is words per clause, third best is clauses per t-unit, and fourth best is words per sentence.

Odell, Lee. "Measuring Changes in Intellectual Processes as One Dimension of Growth in Writing." In Evaluating Writing: Describing, Measuring, Judging. Ed. Charles R. Cooper and Lee Odell. Urbana, Ill.: National Council of Teachers of English, 1977.

A number of contemporary rhetoricians have argued that if students are to improve their writing, they will need to increase their conscious use of certain cognitive and affective processes. By identifying intellectual processes reflected in student writing, we are able to make qualitative distinctions between pieces of writing done in the same mode and for the same audience, diagnose writing problems, and measure growth in writing. Intellectual processes are: focus, physical context, sequence, change contrast, and classification. Each cue can be identified by linguistic cues and by special characteristic vocabulary.

O'Neal, Marcia R., James E. McLean, and George E. McCormick. "The Use of the Syntactic Density Score as an Evaluative Criterion Measure." Tuscaloosa: University of Alabama, 1983. Paper read at the Annual Meeting of the Mid-South Educational Research Association, 16-18 Nov. 1983, Nashville, Tenn. (ERIC Document Reproduction Service No. ED 237 558).

A difficulty associated with the use of Golub and Frederick's Syntactic Density score was the time required in hand tabulation. This drawback was resolved with the development by Kidder of a computer program which calculates a syntactic density score for writing samples. The purpose of this study was to examine the sensitivity of the Syntactic Density computer program as an instrument for measuring changes in language development among high school students. Five items from the program's output were selected for analysis: total number of words, total number of t-units, words per t-unit, total score, and syntactic density score. T-test results show significant gains only for ninth graders and only for three variables: number of words, number of t-units, and total weighted frequency scores. The other three grades showed no significant changes on any of the five variables examined. Nonsignificant results may be accounted for by the time interval from pretest to posttest.

III. RESEARCH RELATED TO CONSTRUCTION OF THE WRITING TASK

A. Procedures for Designing Prompts

Myers, Miles. "Topic Selection." In A Procedure for Writing Assessment and Holistic Scoring. Berkeley, Calif.: University of California, 1980. (ERIC Document Reproduction Service No. ED 193 676).

This booklet describes in detail how to plan for and carry out an all district writing evaluation using the holistic assessment methods developed in the Bay Area Writing Project. To develop a topic for assessment, the BAWP recommends the following steps. First, organize a committee to adopt assessment objectives. Next, match the topic to objectives, selecting mode and rhetorical situations. Then, field test the topic, having students and teachers write on it. Finally, review the topic for problems of unfocused creativity, special knowledge, open-ended form, and grade level differences. Revise the topic if necessary. Never give a prompt that can be answered with yes or no.

Ruth, Leo. "Study 1: Designing Prompts for Holistic Writing Assessments: Knowledge from Theory, Research and Practice." In Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment. Final Report. Berkeley, Calif.: University of California, 1982. (ERIC Document Reproduction Service No. ED 230 576).

This study collected current knowledge from psychometrics, professional wisdom, discourse theory, writing research, and selected practices to find guides to designing effective writing prompts. The findings were: 1) Topics that encourage introspective or autobiographical writing are both motivational and intentional. 2) Instructions are useful, but too many can become constraints. 3) The "describe" trap can be avoided by substituting the instruction "tell about . . ." 4) A very helpful constraint is a suggestion to "focus" on one aspect of the subject. 5) Allowing the student to pick a form may cause a lowered rating, as raters may prefer one form over another. 6) Hypothetical situations make good subjects for writing; however, the need to use the conditional or subjunctive verb phrases can present problems for some. 7) Problems relating to differences in prior knowledge among students can be avoided by providing a body of relevant texts or a literary passage to be commented on.

B. Inclusion of Content in Prompts

Baker, Eva L. and Edys Quellmalz. "Study No. 2, Effects of Topic Familiarity." In Results of Pilot Studies: Effects of Variations in Writing Task Stimuli on the Analysis of Student Writing Performance. Studies in Measurement and Methodology. Work Unit 1: Design and Use of Tests. Los Angeles: University of California, 1979. (ERIC Document Reproduction Service No. ED 213 728).

This pilot study was planned to determine the feasibility of controlling topic knowledge through specific curricular treatment distinctions and to investigate the likely areas of impact of knowledge upon the essay. The pilot study presented a discrete, three-week curriculum-based package on energy to one group of eighth grade students while a comparable group received no such treatment. Both groups received the curriculum-based posttest and were asked to formulate an essay response to a curriculum-based question on energy sources. Analyses compared the two groups on posttest scores and each of the three essay contest measures. The two groups were significantly different in their posttest means: instructed group 83.2% correct, uninstructed group 60.4% correct. The instructed group averaged 172.1 words per essay; the uninstructed group 135.2 words. Between group difference on number of essay facts was 20.1 for the instructed group and 13.1 for the uninstructed group.

Langer, Judith A. "Effects of Topic Knowledge on the Quality and Coherence of Informational Writing." Berkeley: University of California, 1983. (ERIC Document Reproduction Service No. ED 234 418).

This study examined the effect of topic-specific background knowledge on the general quality and local coherence of student writing. Ninety-seven tenth grade students were assigned two writing tasks on specific topics at two points during the semester. Langer's measure of topic-specific knowledge was administered prior to each writing task. Each student's writing sample was then scored on five separate measures: 1) overall quality, 2) coherence, 3) syntactic complexity, 4) audience, and 5) function. The data from the pretest and writing evaluations clearly suggested a strong and consistent relationship between topic-specific background knowledge and the quality of student writing. There is evidence that different kinds of knowledge are predictive of success in different writing tasks. When the assignment calls for a simple reiteration of facts, or elaborations of a given idea, a large amount of unintegrated information will suffice. However, when the student is required to present a thesis, analyze and defend it, the amount of highly organized knowledge, as opposed to simple fluency, will determine success.

Baker, Eva L. and Edys Quellmalz. "Study No. 1, Effects of Writing Prompt Modality on Writing Performance." In Results of Pilot Studies: Effects on Variations in Writing Task Stimuli on the Analysis of Student Writing Performance. Studies in Measurement and Methodology. Work Unit 1: Design and Use of Tests. Los Angeles: University of California, 1979. (ERIC Document Reproduction Service No. ED 213 728).

Thirty eighth grade students wrote an expository or narrative essay in response to an entirely written description of the topic or to a pictorial description of a topic. The written prompt provided minimum cues about the nature of the information desired. The picture prompt portrayed most information necessary for the report requested. T-unit analysis and the CSE expository scale were used to assess each essay. The data suggested that the prompt modalities influenced the form and content of student essays. Essay scores did not differ significantly on the General Impression or Total Scores. However, picture essay scores were significantly higher than written prompt essay scores in the subscales for Essay Organization and Support. The groups did not differ significantly from each other on the total number of words in their essays nor in the number or length of t-units.

C. Prompt Specification of Mode and Audience

Quellmalz, Edys and Frank Cappell. "Defining Writing Domains: Effects of Discourse and Response Mode." Los Angeles: University of California, 1979. (ERIC Document Reproduction Service No. ED 212 661).

The purpose of this study was to examine the stability of measures of student writing performance across types of discourse (genres) and across response modes (multiple choice, constructed response: single paragraph and full length essay). Average or above average eleventh and twelfth graders wrote either expository or narrative paragraphs and essays and took a multiple choice test constructed to sample knowledge of genres. Ratings of the essays and paragraphs on the CSE analytic scale and scores on the objective test provided the basis for the comparison. Results indicated that the relationship between a student's scores on essays written in the same mode of discourse is stronger than the correlation between two essays in different discourse modes. On all five subscales and on total essay and paragraph scores, narrative ratings were lower than expository ratings. On the objective test students answered focus/main idea and support questions similarly well for both expository and narrative passages. On organization questions, however, students had more difficulty with narrative organization.

Prater, Doris. "The Effects of Writing Task on Mastery Classification of High School Students." Clear Lake City, Texas: University of Houston, 1982. Paper read at Annual Meeting of Southwest Educational Research Association, Feb. 1982, Austin, Texas. (ERIC Document Reproduction Service No. ED 215 367).

Eighty tenth grade students, who had been classified as masters or non-masters of writing based on their scores on the writing section of the Texas Assessment of Basic Skills, were assigned three writing tasks, expressive, explanatory, and persuasive essays. Results of the comparison of writing performance across modes of discourse showed that students' expressive writing received the highest rating, followed closely by persuasive writing. Explanatory writing was judged to be the poorest writing overall. The main effect for mastery or non-mastery classification of subjects was not statistically significant. Thirty per-cent of masters and 27% of non-masters varied in their status depending upon the essay used for classification.

Plasse, Lorraine A. "The Influence of Audience on the Assessment of Student Writing." Springfield, Massachusetts: Springfield Public Schools, 1982. (ERIC Document Reproduction Service No. ED 229 760).

To determine the extent of influence that a reader's perspective as a member of a specific audience has on the assessment of twelfth grade student writing, a study examined the holistic judgment made by four different types of writing evaluators on 40 different letters, each of which was written to one of four audience types: student peers, parents, teachers, and business people. A statistical analysis of the data showed that the audience to whom students wrote was a significant factor in the scoring of the letters. The mean holistic ratings of letters written for the business and teacher audiences were significantly greater than were the mean ratings of letters written for either the student or parent audiences. All four types of raters preferred letters written to teacher and business audiences. Contrary to assumptions of rhetoricians who suggest that readers will respond more favorably to texts that specifically address them than to those that do not, writing evaluators in this study did not necessarily prefer the letters that were written to the audience types which they represented.

D. Presentation and Interpretation of Prompt

Caroll, Karen and Sandra Murphy. "A Study of the Construction of the Meanings of a Writing Prompt by Its Authors, the Student Writers, and the Raters." In Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment. Final Report. James Gray and Others. Berkeley, Calif.: University of California, 1982. (ERIC Document Reproduction Service No. ED 230 576).

This study investigated the problem of student misunderstanding of teacher intention in comprehending the writing prompt. Students were interviewed and recorded concerning their interpretation of the prompt prior to writing and immediately after writing. Teachers were asked during construction of the prompt to verbalize their expectations of students' responses. Data in the study indicated that expectations of evaluators differed from those of student-writers. One major difference in prompt-perception was that the prompt authors lent weight to all parts of the prompt, while some of the student-writers focused on only a few selected aspects of the prompt. Interpretations of prompt segments influenced rhetorical strategies adopted by students in composing a response. Those student-writers who rated the prompt easy to read and understand, and whose interpretation of the prompt agreed with that of the authors, tended to receive the higher holistic scores.

Camp, Gerald. "Study 3: Part II: A Field Test of Eight Prompts: Santa Clara County Writing Assessment." In Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment. Final Report. James Gray and Others. Berkeley, Calif.: University of California, 1982. (ERIC Document Reproduction Service No. ED 230 576).

The goal of this study was to obtain information on the performance on each of the eight prompts developed for use in the Santa Clara County Writing Assessment in the typical county schools identified as low, mid-range, and high. Teacher-raters were asked questions after completing the scoring of the papers. The responses to these questions showed very little agreement among the teacher-raters as to which of the eight prompts were less and which were more successful. Almost all of the prompts were successful in producing the full range of writing quality desirable in such a sampling.

Odell, Lee. "Defining and Assessing Competence in Writing." In The Nature and Measurement of Competency in English. Ed. Charles R. Cooper. Urbana, Ill.: National Council of Teachers of English, 1981.

Procedures for evaluating writing must be consistent with our best understanding of writing. Writing skills include discovering ideas or invention and making choices about voice, audience, and purpose. To allow students full use of all components of the writing process during the assessment process, procedures for obtaining an adequate sample of student writing during evaluation should include: 1) giving students enough time to explore their topic, consider their audience and purpose, and revise and edit their work, 2) having them do more than one kind of writing, 3) providing them with information about the audience and purpose for which a given piece of writing is intended, 4) basing our judgments on an adequate amount of student writing (one sample is not reliable), and 5) after testing, assessing the demands of our writing assignments in terms of products and grades received.

Myers, Miles. "Directions." In Procedure for Writing Assessment and Holistic Scoring. Berkeley, California: University of California, 1980. (ERIC Document Reproduction Service No. ED 193 676).

This booklet summarizes the best way to give directions when assessing writing, based on the Bay Area Writing Project experience. It was recommended that: 1) the evaluation time be arranged to fit the class schedule and the attention span of the students, 2) the number codes for information desired be put on the samples in advance, 3) coding directions for students be prepared in advance, 4) directions for teachers stress the importance of giving encouragement but no assistance with the content of the prompt, 5) after distribution, teachers write the prompt clearly on the blackboard, 6) teachers stress to students the importance of dark copies and some kind of response, 7) directions be varied to fit the needs of different grade levels, allowing K-3 students to draw and dictate or draw and write.

IV. FACTORS INFLUENCING EVALUATORS OF WRITING SAMPLES

A. Evidence of Differences in Rating

Baker, Eva L. and Edys Quellmalz. "Study No. 3, Effects of Topic, Sample and Rater Group Membership on the Stability of Scoring Criteria Application." In Results of Pilot Studies: Effects of Variations in Writing Task Stimuli on the Analysis of Student Writing Performance. Studies in Measurement and Methodology. Work Unit 1: Design and Use of Tests. Los Angeles: University of California, 1979.

This study examined the influence of three potential sources of error in writing evaluation: composition of rating group membership, variations in the quality of student writing sample, and changes in topic. Rating context variables were level of rater experience with the scale and training group membership. Scores were compared on essays written on three different narrative topics by high school seniors of varied academic achievement. Two rater pairs scored essays in both studies A and B. Examinees in study A were drawn from average and above average high school English classes. Students in study B were high school students admitted to the Univ. of Calif. Raters used the CSE scale. Results showed that topic assignment and quality range did seem to affect scoring. Also, rater's scale experience and training group membership did seem to yield different essay scores.

B. Factors Relating to the Writing Sample

Breland, Hunter N. and Robert J. Jones. "Perceptions of Writing Skills." Written Communication 1.1 (Jan. 1984): 101-119. (ERIC No. EJ 293 034).

This study attempted to discover what criteria raters use when making holistic judgments of brief, impromptu essays and if raters judge such essays in the same way that they thought they did. Holistic scores assigned to a sample group of essays were compared with detailed analysis of the same essays made subsequent to the original scoring. For the second question, English professors were asked, prior to their actual analysis of essays, to state what they believed to be the most important characteristics of these essays. Results showed that overall organization, use of supporting materials, and noteworthy ideas were the most important factors in reader judgments. Syntactic and lexical characteristics were less important. Raters thought that they were influenced by overall organization, statement of thesis, supporting material, noteworthy ideas, and mechanics. Implications for instruction suggest that greater instructional emphasis is needed on discourse skills rather than mechanics, since teachers value content and organization more than mechanics.

Harris, Winifred Hall. "Teacher Response to Student Writing: A Study of the Response Patterns of High School English Teachers to Determine the Basis for Teacher Judgment of Student Writing." Research in Teaching English 11 (Fall 1977): 175-85.

The study tried to discover the response patterns of 36 high school English teachers to 12 selected high school writing samples. The teachers were asked to mark the themes as they normally did. Next they ranked the papers from best to worst and filled out a questionnaire designed to elicit their reasons for grading. The response patterns which emerged indicated that there were some discrepancies between theory and practice when student compositions were actually evaluated. Although the teachers' responses revealed a slight preference for giving more weight to content and organization, when they rated the samples they were strongly influenced by errors in mechanics and usage. If a student composition was deficient in this category, despite the fact that it exhibited merit in content and organization, the paper was likely to be ranked lower by almost one-half of the teachers than a theme which exhibited merit in mechanics and usage but was deficient in content and organization.

Freedman, Sarah Warshauer. "Some Reasons for the Grades We Give Compositions." English Journal 71 (Nov. 1982): 86-89. (ERIC No. EJ 269 776).

The purpose of this study was to see if English teachers value mechanics above development and organization, or valued development most. The research took a selected set of previously scored average quality expository essays written by college freshmen and had them rewritten to be weaker or stronger in 1) development or content, 2) organization, 3) sentence structure, 4) mechanics. Results of the study showed that teachers gave significantly higher scores to papers rewritten to be strong in development, organization, and mechanics than they did to papers rewritten to be weak in those areas. The development rewriting was most influential, the organizational rewriting next, and mechanics third. Teachers apparently felt that if the organization was strong, then the quality of the mechanics mattered even more than it did on its own and the quality of the sentence structure mattered as well. If the discourse level qualities were weak, the paper deserved a low score, and strong sentence structure and mechanics could not redeem such papers.

Sloan, Charles A. and Iris McGinnis. "The Effect of Handwriting on Teachers Grading of High School Essays." De Kalb, Illinois: Northern Illinois University, 1978. (ERIC Document Reproduction Service No. ED 220 836).

This study was conducted to investigate the effects of handwriting on the grades assigned to essays. Papers were graded rapidly and holistically. Good handwriting was defined as that done by five experts in the Palmer Handwriting Method. The essays were 200 to 250 word themes written by 500 ninth graders. Forty-five were randomly selected and rewritten by the Palmer experts. The copies, plus the original student-written themes were scored by a team of 12 teachers trained in the Diederich holistic method. A t-test for correlated data was conducted to analyze the difference between scores on the students' original papers and papers rewritten by the experts. The results indicated that the papers written in expert handwriting were evaluated by the teachers as significantly better papers. The results also indicated that the scores of the upper two-thirds of the papers tended to improve more than those of the lower third.

C. Factors Relating to Teacher Expectations of the Writer

Diederich, Paul Bernard. "The Effect of Bias." In Measuring Growth in English. Urbana, Ill.: National Council of Teachers of English, 1974.

Teachers should avoid bias either for or against particular students, the views expressed, the way of writing (ornamented or plain, length or brevity). Some teachers react strongly to particular errors. Bias appears most obvious when teacher grades the papers of his own students. Even when the paper of a given student surprises or disappoints the grader, the teacher is likely to change too little. In an experiment conducted by Rosner and Diederich of ETS to test methods of teaching writing, teacher-raters asked to know grade and level of papers they were grading. Grade and level (regular/honors) were stamped on the papers in an alternate, random fashion. The study concluded that the papers stamped honors averaged almost one grade-point higher than the other copies of the very same papers that were stamped regular.

Gronland, Norman E. "Measuring Complex Achievement: The Essay Test." In Measurement and Evaluation in Teaching. New York: Macmillan Publishing Co., Inc., 1981.

Teachers may score essays unreliably because of the halo effect. The halo effect in a long essay test may be caused by a certain general impression made by the first few answers to the questions. These answers color rater judgments of remaining answers. The halo effect is also caused by the general impression a teacher forms about each pupil during classroom teaching. Recommended ways to control the halo effect are to evaluate all answers to each question at one time, and to evaluate without looking at the pupils' names.

Wade, Barrie. "Responses to Written Work: The Possibilities of Utilizing Pupils Perceptions." Educational Review 30 (June 1978): 149-58. (ERIC No. EJ 188 175).

This article surveys research about the effect of students' personal qualities on teacher assessment of their writing. The study showed that students' first names created expectations and influenced the results of evaluations. Another study suggested that physical appearance of a student influenced a marker. Some examiners were influenced by the mode of presentation as well as the attractiveness of the writer. Speech accents could create a favorable or unfavorable reaction in teachers. Handwriting strongly influenced assessment of the quality of essays.

V. RESEARCH RELATING TO RESPONSE CRITERIA FOR WRITING EVALUATION

A. Method of Administering Response Criteria

Diederich, Paul Bernard. "Measuring Improvement in Writing." In Measuring Growth in English. Urbana, Ill.: National Council of Teachers of English, 1974.

Bias in grading test papers is avoided by the following procedure of evaluating writing. All students in a span of three or four grades write a paper on the same topic on the same day. Each student gives his paper a six digit number and does not sign his name. For evaluation the papers are arranged in order of these self-chosen numbers, with three or four grades scrambled together. Each reader records his grades and comments on a separate work sheet to avoid biasing the judgment of the second reader. Then the papers are turned over to another teacher for a second, independent rating. After two evaluations, papers which differ in grades by more than one grade point (or stanine) will be reconciled by a different, very experienced rater rereading the paper and substituting a new grade for the one that was farthest from it.

Myers, Miles. "Scoring." In Procedure for Writing Assessment and Holistic Scoring. Berkeley, Calif.: University of California, 1980. (ERIC Document Reproduction Service No. ED 193 676).

The following checklist follows the recommended scoring procedures used in experimental research of the BAWP. 1) Select table leaders (expert markers) and estimate the number of teachers and days needed for the reading. 2) Table leaders select anchor papers to define each score from the sample. One problem will be the selection of the anchors which define the difference between upper-half and lower-half papers. 3) Table leaders identify traits of scoring categories. 4) Table leaders score papers to test reliability of anchors. 5) Table leaders identify papers that focus on a particular problem, especially the relative weights of fluency and coherence. 6) Begin complete evaluation by having readers score a set of anchor papers, one for each scoring category. 7) Have readers score a second set of anchors with more papers in the middle categories. 8) Have readers score special anchors which focus on any problem trends noticed. Splitter papers (papers widely distributed in categories) are usually put in the lower half. 9) Have papers read twice by different readers and a third time if the two scores differ by more than one point.

Nyberg, Verner R. and Adell N. Nyberg. "Alberta Essay Scales: Models." Edmonton, Canada: University of Alberta, 1982. (ERIC Document Reproduction Service No. ED 220 483).

This document gave directions to raters in the use of the Alberta Essay Scale Models. Scores for each model represented stanines 2, 4, 6, 8. A separate set of essay models was included for both mechanics and style/content. The following routine was suggested for scoring the essays: 1) Read the essay to be graded, giving attention to style and content only. 2) Form a general impression of how well the writer organized his thoughts and communicated them. Ignore all mechanical errors. 3) Note strengths and weaknesses, bearing in mind the style/content variables. 4) Rater should match his impression of the essay with the model that came nearest to creating the same impression, then award the appropriate scale score for style/content. 5) Reread the essay placing a mark on each mechanical error discovered. 6) Count the errors and express the total as a rate of number of errors per 100 words. 7) Find the model having an error rate nearest to the rate for the essay to be graded and award the corresponding stanine value.

B. Validity of Atomistic Response Criteria

Chew, Charles R. "A Study to Determine the Relationship Between Indexes of Maturity and a Quality Grade in a Written Examination." Albany: New York State Education Dept., 1978. (ERIC Document Reproduction Service No. ED 178 914).

The purpose of this study was to determine the maturity of students' writing which was done for the 1977 New York Regents Comprehensive Examination in English, to discover if a relationship existed between a teacher's holistic evaluative grade and the maturity of writing as determined by Kellogg W. Hunt's structural analysis, and to compare the maturity of writing done for the examination with Hunt's findings of syntactic maturity at three grade levels. The conclusion of the study was that no statement could be made at this point to suggest that evidence of certain indexes of syntactical development in a paper guaranteed top quality of writing or the lack of it. The mean t-unit length of the top three quality groups corresponded to the length Hunt prescribed for grade 12. The papers with the longest t-unit length were among those receiving the highest grades. Nearly one-third of the papers had t-unit lengths less than the length that Hunt prescribed for grade 8, but only 23% received the lowest quality grade. Mean clause lengths were longer for the top quality groups.

Tierney, Robert J. and James H. Mosethal. "Cohesion and Textual Coherence." Research in the Teaching of English 17.3 (Oct. 1983): 215-29. (ERIC No. EJ 289 567).

This study explored the relationship between Halliday and Hason's system for analyzing cohesion and instructors' rankings of textual coherence. Essays of twelfth grade students were analyzed for various cohesive ties and were also rated holistically with respect to clarity of expression and general coherence. The results of the comparison argue against using cohesion analysis as an index of textual coherence. There appears to be no causal relationship between proportional measures of cohesive ties and holistic coherence rankings. There was very little variation in cohesive patterning among essays written on the same topic.

Stotsky, Sandra. "Examining Word Use to Assess Growth in Essay Writing." This essay is the second revision of a paper presented at the 6th Annual Conference on Language Development, Oct., 1981, Boston. (ERIC Document Reproduction Service No. ED 211 980).

Twenty tenth grade students participated in a study designed to identify differences in the ways developing writers use words to create written texts. Essays written by the students were holistically evaluated and then analyzed for the types and number of lexical cohesive ties used and the nature of the grammatical subject, or focus of each clause. The analysis of lexical cohesion in the nine high-rated and eleven low-rated essays showed large differences. In the low-rated essays, the total number of lexical ties constituted 15% of the total number of words, the total number of words entering into lexical ties constituted 25% of the total number of words. In the high-rated essays, the total number of lexical ties constituted 44% of the total number of words. The analysis of the grammatical subjects of clauses also revealed large differences between the two groups of writers. Of the total number of 121 clauses in the low-rated essays, 81% were pronominal in nature. In contrast, of the 154 subjects of clauses in the high-rated essays, only 51% were pronominal. Writers of the high-rated essays were clearly using words in different ways than the writers of the low-rated papers.

C. Comparisons of Other Direct and Indirect Response Criteria

Winters, Lynn. "The Effects of Differing Response Criteria on the Assessment of Writing Competence." Los Angeles: University of California, 1978. (ERIC Document Reproduction Service No. ED 212 659).

The purpose of this study was to investigate the relative validities of four essay scoring systems. Each rater was trained in two of the four scoring systems: General Impression Scoring (GI), Diederich Expository Scale (DES), Center for the Study of Evaluation Analytic Scale (CSE), or T-Unit Analysis (T). Writing samples were randomly selected from four composition classes: low performance high school writers, high performance high school writers, low performance college writers, and high performance college writers. All four scoring systems had high reliability. Three of the systems (GI, DES, CSE) were associated with some differences in group performance. T-unit was not, making it virtually useless for classification purposes. The classification accuracy of the systems was questionable in distributing all but the high school low group. The DES contributed most to group separation when combined with either CSE or GI. Both DES and CSE contributed more to discrimination than GI. The use of the three scoring systems produced more accurate classification results than using scores on only DES and CSE.

Smith, Laura Spooner. "Investigation of Writing Assessment Strategies. Studies in Measurement and Methodology. Work Unit 1: Design and Use of Tests." Los Angeles: University of California, 1978. (ERIC Document Reproduction Service No. ED 213 727).

The purpose of this study was to explore the relationships among production form of response in writing samples, judged by analytic and impressionistic scoring criteria, and selection form of response, an objective test developed to measure skills presumably related to the production of written discourse. The samples were 128 students from six different English classes from three high schools in the Los Angeles area with varied abilities. Analyses performed on the data found that scores yielded by the impressionistic and analytic rating strategies were highly related. The six subscales comprising the analytic rating scale were highly intercorrelated. Of the six analytic subscales, mechanics and support proved to be most strongly associated with impressionistic rating scores. The objective measure total score proved to be moderately related to both analytic and impressionistic total scale scores.

Veal, Ramon L. and Sally Ann Hudson. "Direct and Indirect Measures for Large-Scale Evaluations of Writing." Research in the Teaching of English 17 (Oct. 1983): 290-96. (ERIC No. EJ 289 573).

The purpose of this study was to present reliability, validity, and cost comparisons of several direct and indirect measures of writing. The direct measures were holistic, analytic, primary trait, and mechanics counts scoring of actual writing samples. The indirect measures included the language arts items of the following objective tests: California Achievement Test, Iowa Test of Basic Skills, Tests of Achievement and Proficiency, and Writing Proficiency Test. Subjects were 100 tenth graders randomly selected from 24 Georgia high schools. Each participating school received two tests of different types. Inter-rater reliability for the holistic scorings ranged from .69 to .76, analytic ratings .64. The correlation between the two primary trait raters was .72. There was a minor correlation between papers rated holistically and by mechanics counts. Costs for scoring papers were: objective tests $.53 per student, holistic $.75, analytic $2.37, primary trait $.58, and mechanics counts $1.06. The ITBS showed the strongest correlation with the holistic ratings: .50 to .60. Thus, in view of validity, reliability, and cost, holistically rated writing samples were recommended.

Bauer, Barbara Ann. "A Study of the Reliabilities and the Cost-Efficiencies of Three Methods of Assessments for Writing Ability." Champaign: University of Illinois, 1981. (ERIC Document Reproduction Service No. ED 216 357).

A group of raters scored 120 essays written by 17 year olds, taken from the narrative-descriptive mode of discourse in the 1973-74 NAEP writing assessment. The raters used the analytic, holistic, and primary trait scoring methods. The inter-rater reliability estimates were highest for the analytic method, and lowest for the primary trait method (.874-.633). Intra-rater reliability was highest for raters using the analytic method (analytic average .900, primary trait average .836, holistic average .785). The average and composite correlation coefficients were highest for the analytic scoring method, followed by the primary trait method, and finally the holistic scoring method. Cost was described in terms of rater-training time and average rating time per essay. The training times were analytic method 114 min., holistic method 52 min. and primary trait method 44 min. The average time needed to grade a single essay was analytic 4.14 min., holistic .96 min., and primary trait 1.08 min. It was concluded that the analytic scoring method was the most reliable and the holistic method was the most cost-efficient.

Koffler, Stephen. "Assessing Students' Writing Skills: A Comparison of Direct and Indirect Methods." Trenton, New Jersey: New Jersey State Dept. of Education, 1984. Paper read at the annual conference of the American Educational Research Association, April 1, 1984, New Orleans. (ERIC Document Reproduction Service No. ED 242 771).

This research examined the results from direct and indirect writing assessments to determine the most effective method of discrimination. It sampled the 94,000 ninth graders who took both the objective test of writing ability and wrote an essay which was holistically scored. Following the administration of the test, teachers predicted whether they thought that each student would master or non-master or borderline the test they had just completed. When the results were tallied, teacher judgments held. Those labeled master had higher scores than those labeled non-master. When the parts of the multiple choice test were broken down, the sentence-structure cluster of the test was a better predictor or discriminator of mastery than the essay. The essay results predicted false masters 44% of the time. The multiple choice questions predicted false masters 30.7% of the time. Statistical analysis indicated the indirect assessment provided better means of discrimination between competent and incompetent writing.

SUMMARY OF RESEARCH

I. Development of Writing Skills

Research indicated that the early writing of children developed from the speech they heard and participated in in their own homes. Fadiman (1979), Macintosh (1953), Britton (1970), and Bartlett (1981) emphasized that children developed incentives to write when ample purposeful writing activities were presented to them, with freedom to make mistakes without adult disapproval. Britton also traced the developing use of discourse mode in young writers. He found that young children began writing in expressive modes and progressed to more formalized participant and spectator uses of language. The adolescent's tendency to be alienated from adults affected his use of language and his valuing of writing (Fadiman, 1979 and Britton, 1970).

Britton (1975) analyzed the development of audience and function in writing done in school by British adolescents. The teacher was the most common audience for students' writing, beginning with student-teacher dialogue writing in early years and moving towards an increase of teacher-as-examiner writing in the later school years. When analyzing the same school writing in terms of mode, Britton found that the transactional or participant categories predominated with large amounts of analogic writing present throughout the secondary school years. In the last

school years, only small amounts of writing on the more abstract levels was done. Bartlett (1981) found that children's knowledge of features of the particular kinds of discourse developed slowly.

Linguists traced the development of writing skills toward the goal of syntactic maturity which Odell (1979) described as the time when the writer acknowledges the differences between himself and his audience and is aware of the limitations of his own point-of-view. Hunt (1977) described the development of syntactic maturity in terms of student ability to write more complex t-units, achieve coordination of objects, use appositives, remove adjectives from the nominative position, use verbs as participial phrases, and make syntactic category shifts. In a series of studies of fourth through ninth grade children's abilities to revise their own and others' writing, Bartlett (1981) found that children of these age groups had difficulty detecting textual inadequacies in their own writing and incorporating new information into revision of their work. Benton (1984) found that good youthful writers had better ability to hold many items in short term memory and simultaneously manipulate them rapidly, than did poor writers.

II. Descriptions of Methods of Writing Evaluation

There are two distinct styles of writing evaluation—

the direct and the indirect methods. The direct method of evaluation asks students to produce a piece of writing, which can then be evaluated in a number of different manners. The indirect method of evaluation asks students to select responses on an objective test to prove facility in certain processes related to writing skills. According to Hall (1981) and Braddock (1976), these standardized writing achievement tests measured subcomponents of writing, such as vocabulary knowledge, grammatical usage, mechanics, topic sentences, sentence sequence and elimination of irrelevant sentences. The theory behind this type of testing was that if children could find textual anomalies in a printed text, they should be able to avoid the same problems when they actually produced writing of their own.

The direct methods of evaluating writing can be classified into four types: holistic, analytic, primary trait, and atomistic. Holistic scoring is the name for the process used when the rater quickly reads an essay and responds to it as a whole. The evaluator will give it a numerical score, which represents the overall quality of the essay. Holistic scoring can be also called general impression marking (GI) or an impressionistic scale. Studies which described this kind of evaluation included: Alloways' (1981) description of the Holistic Scoring plan for the New Jersey Writing Project, Bamberg's (1984) reassessment of NAEP essays for global coherence, Chew's (1981) description of holistic assessment in the New York Regent's Competency Test,

the Alberta Essay Scale Models (1982) (which were developed analytically but are used in a holistic manner), and Spencer's (1979) explanation of In-School Folio Assessment in Scottish Secondary Schools.

Another direct measure, primary trait scoring, is very similar to holistic scoring in that it is done very rapidly. However, primary trait scoring is genre-specific. A primary trait rubric will be designed for a particular mode of writing, such as exposition, narration, or persuasion. A rater using primary trait scoring notes the presence or absence of particular significant features of the chosen rhetorical mode. The score is a tally of those features. Lloyd-Jones (1977) described the method in general and gave specific examples of scoring guides for narration and persuasion. Rabianski (1977) devised an informative scale which is actually a primary trait schema for scoring writing that contains generalizations supported by elaboration.

Many examples of analytic scales are given in this study. A rater using an analytic scale examines and scores separately different aspects of the written script and then gives a rating which is a tally of grades for separate parts. Some scales include as one of the parts a general impression or overall effect score. Perhaps the oldest scale is the Diederich (1974) Expository Scale (DES). The CSE Expository and Narrative Scales (Burry and Quellmalz, 1983) were developed at the Center for the Study of Evalua-

tion at UCLA. Malgady and Borcher (1979) devised an analytic scale to measure creativity which correlated positively with holistic evaluations of creative writing.

Atomistic procedures for evaluation require that the rater count explicit features of text to arrive at a measure of quality. These measures seek to detect the presence or absence of such features as cohesive ties, t-units, number of words, certain kinds of vocabulary, certain intellectual processes. Hunt (1970) defined the t-unit, a language measurement that can be used to evaluate the maturity of a student's writing. A t-unit is a single main clause (or independent clause) plus whatever other subordinate clauses or nonclauses are attached to, or embedded within, that one main clause. O'Neal (1983) described a computer program used to calculate syntactic density scores. This program counted total number of words, total number of t-units, total number of words per t-unit—all of which combined to provide a syntactic density score. Odell (1977) described a method of evaluation which measured linguistic cues for changes in intellectual processes.

III. Research Related to Construction and Use of Writing Task or Prompt

The writing task or prompt is the assignment for writing given to the students in an assessment situation. Evaluation research tells the educator much about prompt selection and content, discourse mode and audience addressed,

student interpretation of the prompt, and method of presentation of the writing task. Myers' (1980) research suggested that a prompt should be carefully tied to curricular and measurement objectives. The prompt should then be field tested to assure that those objectives were indeed attained. Ruth (1982) compiled a list of suggestions for precise wording of the writing prompt. Two research studies by Langer (1983) and Baker and Quellmalz (1979, Study No. 2) demonstrated that students who were given information about the subject prior to assessment wrote better quality essays than students who were not. In a different study, Baker and Quellmalz (1979, Study No. 1) also discovered that students wrote better essays from a visual prompt which included a great deal of information than from a written prompt which included very little information.

Specification of audience and mode in a writing prompt can influence student scores. A study conducted by Quellmalz and Capell (1979) found that students writing narrative essays received lower ratings than students writing expository essays. Prater (1982) found that expressive writing samples received the highest ratings, followed closely by persuasive, with explanatory samples being rated lowest. Regarding audience addressed, Plasse's (1982) study found that highest holistic assessment scores went to letters addressed to businessmen or teacher audiences as opposed to being addressed to peers and parents.

Authors and student responders often interpret a writing prompt very differently, according to research conducted by Caroll and Murphy (1982). Furthermore, those students who interpreted a writing task in a manner similar to its authors' intent usually received higher holistic scores on their writing. Both Odell (1981) and Myers (1982) agreed that a writing task should be presented to a group of students in a precise and purposeful manner.

IV. Factors Influencing Evaluators of Writing Samples

Teachers and others who evaluate writing are subject to influences which may bias their assessment of the script. The fact that rater pairs may differ was shown in a Baker and Quellmalz study (1979, Study No. 3). Certain aspects of the papers themselves influence the rater, but exactly what these factors are seems to vary with the study. Breland and Jones (1984) found that while teachers thought they would take mechanics into consideration when grading, they actually put most weight on organization, support, and ideas. On the contrary, both the Freedman (1982) and the Harris (1977) studies concluded that teachers did include mechanics as a major factor in evaluating essays. Sloan (1978) showed that essays written in a superior handwriting received higher grades than the same papers written in average or poor handwriting. Teacher expectations of students might also influence their evaluations. Diederich (1974) found that papers labeled "honors" were

graded higher than papers labeled "regular." Gronlund (1981) described a "halo effect" operating for or against certain students because of their previous performances. Personal qualities of the students such as appearance, name, or speech accent might also affect grades, according to Wade (1978).

V. Research Relating to Response Criteria for Writing Evaluation

The most effective procedures for evaluating large numbers of writing samples have been laid out by experts in the field. Diederich (1974) advised coding each paper by number and having it evaluated by two raters. After two samples from each writer were evaluated, papers with wide rating discrepancies would be reconciled by having a different more experienced evaluator rate them again and replace the most discrepant grade. Myers (1980) of the BAWP first selected table leaders who picked anchor papers. Anchor papers then became models for regular graders. Again each paper was rated two times. Nyberg (1982) described the procedure of matching essays to similar models which was used with the Alberta Essay Scales.

The research about the validity of different types of atomistic response criteria produced conflicting results. In a comparison of papers rated both holistically and with atomistic measures, Chew (1978) found that the atomistic measures such as t-unit length, clause length, and ratio of

clauses to t-unit did not guarantee either top quality writing or the lack of it. The research results from studies using cohesive measures of evaluation were contradictory. Tierney and Mosethal's study (1983) showed no causal relationship between proportional measures of cohesive ties and a holistic rating of coherence. However, when using a different system of measuring lexical cohesion specifically designed for high school students' writing, Stotsky (1981) found that the atomistic measure did distinguish significantly between high and low rated essays.

Comparison of other direct and indirect response criteria brought varied results. Winters' (1978) comparison of general impression grading (GI), Diederich expository scale (DES), the CSE analytic scale, and t-unit analysis discovered that each system produced different results. The t-unit proved useless for showing quality differences among papers and the other three measures only distinguished the low group. Smith's (1978) comparison of an indirect measure (an objective test specially constructed to reflect skills listed in the analytic scale), an analytic scale, and an impressionistic scale found that the impressionistic and analytic ratings were highly correlated with each other and that the objective test was moderately related to both. Veal and Hudson (1983), in a comparison of validity, reliability and cost of holistic, analytic, primary trait, mechanics count scoring, and objective tests found that holistic scoring was acceptably valid, reliable, and most econo-

mical. Of many indirect measures tested, the Iowa Test of Basic Skills was the one showing the strongest correlation with holistic ratings. In another study Bauer (1981) found analytic scoring the most reliable and holistic scoring the most cost effective in assessing large numbers of students. Koffler's (1984) study of a large number of writing samples produced by New Jersey students, using teacher judgments of mastery or non-mastery status as a criterion reference, found the indirect method—the objective test—a better predictor than the holistically scored essay.

CONCLUSIONS

The conclusions to this study fall into two general categories: those relating to the framing of the evaluation task or instrument, and those relating to the response criteria or the methods of actually assessing writing samples.

The linguistic and psychological treatises examined in this study prove that indirect measures of writing evaluation are not acceptable. Direct methods of writing evaluation are preferred to indirect because Bartlett's (1981) research suggests that young people between grades four through nine have difficulty detecting certain anomalies in their own text, even though they can detect similar anomalies in the texts of others. Furthermore, most students of this age group can detect referential ambiguities, but seem to lack the mental flexibility to supply the missing referential information when rewriting their own texts. Bartlett's findings cast doubt upon the relationship of the editing activities found in most standardized tests of writing to the actual acts of composition and rewriting. Bartlett's research further suggests that in order to establish validity, each item on a standardized test ought to be developmentally referenced, to assure that young people aren't asked to find or correct items which are beyond their level of cognitive development.

Testing children's writing by having them write, the direct method, takes cognizance of current psychological

theories of writing skill development. Macintosh (1953), Fadiman (1979), and Britton (1970) all stated that children learn to write by means of purposeful assignments. Testing children's writing skills directly will encourage classroom teachers to make children write often. Testing writing skills with multiple-choice tests will encourage classroom teachers to make children practice minute aspects of the writing task rather than having them write complete pieces.

Certain indirect measures have been shown to have a moderate correlation with direct measures in the studies of Smith (1978) and Veal and Hudson (1983). Those indirect measures which have been constructed to contain skills objectives found in analytic scales seem to correlate most highly. Because of the correlations, those types of indirect measures may be used in large-scale assessments as criterion measures, but they should not comprise the totality of any writing assessment.

Research abstracted in this study about the evaluation task presents a conclusive picture of how to formulate the writing task. Much research addresses the specifications for wording the actual writing assignment given to the students, the task. The writing assignment must be expressed simply, using unambiguous vocabulary and sufficient detail to precisely define expectations (Ruth, 1982). Specification of discourse mode for the writing task should be appropriate to the developmental level of the students. Writing

tasks framed in the expressive mode are suitable for elementary children. Only grades eleven and twelve have the developmental ability to respond well to the more abstract transactional modes of writing (Britton, 1970, 1975). Task authors must ascertain whether subjects have had classroom experience writing in a particular mode before it is assigned on an evaluation, because the literature shows great scoring variations between modes of discourse (Quellmalz and Capell, 1979). Since the audience to whom the writing task is directed does influence the quality of the results (Plasse, 1982), it is important that the audience type prescribed in the writing task be one which measurement experience demonstrates will produce a product of a quality level acceptable to the evaluators. The administration of the writing task should allow the subject enough time to implement the entire writing process of conception, incubation, and production, particularly for younger students who have limited short-term memory (Benton, 1984, Britton, 1975).

Factors relating to the writing sample and to teacher expectations of the writer which negatively influence raters' or teachers' evaluation of writing samples can be controlled in large-scale writing assessments by following the procedures for conducting an evaluation suggested by Diederich (1974) and Myers (1980). These procedures include using coded papers, rating each paper twice, adjudicating discrepantly graded papers, training the evaluators, and

monitoring graders during the evaluation process.

The research analyzing the different methods of writing evaluation response criteria is suggestive rather than definitive. The research available at this time (Chew, 1978, Tierney and Mosethal, 1983, Stotsky, 1981) suggests that atomistic response criteria such as t-unit analysis and measures of cohesion do not accurately predict overall quality in essay writing. Therefore it is recommended that the use of these methods of writing evaluation be discontinued until research produces a definitive answer.

Comparative studies of other direct response criteria (Winters, 1978 and Bauer, 1981) suggest that analytic methods of assessment are the most reliable and valid. The disadvantage of the analytic scale is its high administrative cost. Since the relatively inexpensive holistic measurement method correlates reasonably well with analytic measures (Smith, 1978), holistic assessment will probably be the pragmatic choice for large scale assessments.

More research needs to be done in the area of writing evaluation response criteria. Of the five studies on response criteria examined in this paper, two had serious flaws. The Koffler (1974) study concluded that a holistically graded essay was not a good discriminator of writing mastery, using the prediction of the student's classroom teacher as the criterion reference. Other studies cited in this paper (Diederich, 1974, Gronlund, 1981, and Wade, 1978) show that teacher opinion can be very biased. There-

fore, Koffler's conclusion about the superiority of indirect measurement instruments is open to question. Winters' study (1978) of differing response criteria concluded that three methods had difficulty classifying any but the lowest group of writers. Her criterion reference was a student classification of subjects into high school low, high school high, college low, and college high groups. The original classification of the college groups was open to question, because the study was conducted during a summer session, when English-as-a-second language students were scheduled into both classes on the basis of a single, indirect test. Thus, if the criterion group classification was open to question, so were the results of Winters' study. With two studies removed from consideration, only three valid ones remain—Smith (1978), Veal and Hudson (1983), and Bauer (1981). This is a small number of studies upon which to base theory for writing competency testing, particularly when the test results may determine high school graduation. Therefore, the conclusions put forth in this paper favoring analytic and holistic evaluation methods must remain tentative until new research is conducted.

RECOMMENDATIONS

It is recommended that:

1. More research studies be conducted to establish the reliability and validity of holistic, analytic,

and primary trait response criteria.

2. The major assessment method in large scale writing assessments be direct measures of writing performance. The analytic method is recommended when obtaining feedback is required, and the holistic method is recommended when economy is the major consideration.
3. Preprofessional education of classroom English teachers include course work in developmental linguistics and psychology to enable them to use a variety of evaluation methods and to make them more sensitive to the relationship between linguistic development of their students and their performance in evaluation tasks.

BIBLIOGRAPHY

Alloway, Evans. "Holistic Scoring Session for the New Jersey Writing Project: A Summary." Trenton, New Jersey: May, 1979. Report prepared by the New Jersey Writing Project, New Jersey State Dept. of Education. (ERIC Document Reproduction Service No. ED 191 069). (26, 49).

Baker, Eva L., and Edys Quellmalz. "Results of Pilot Studies: Effects of Variations in Writing Task Stimuli on the Analysis of Student Writing Performance. Studies in Measurement and Methodology, Work Unit 1: Design and Use of Tests." Los Angeles: University of California, 1979. (ERIC Document Reproduction Service No. ED 213 728). Studies No. 1, 2, and 3 abstracted. (33-34, 38, 52-53).

Bamberg, Betty. "Assessing Coherence: A Reanalysis of Essays Written for the National Assessment of Educational Progress, 1969-1979." Research in the Teaching of English 18 (Oct. 1984): 305-19. (ERIC No. EJ 305 318). (27, 49).

Bartlett, Elsa Jaffe. Learning to Write: Some Cognitive and Linguistic Components. Papers in Applied Linguistics, Linguistics and Literacy Series: 2. Washington D.C.: Center for Applied Linguistics, 1981. Chapters 1-3 abstracted. (7, 13, 19, 21, 47-48, 57).

Bauer, Barbara Ann. "A Study of the Reliabilities and the Cost Efficiencies of Three Methods of Assessment for

Writing Ability." Champaign: University of Illinois, 1981. (ERIC Document Reproduction Service No. ED 216 357). (46, 56, 60-61).

Benton, S.L., Robert G. Kraft, John A. Glover, and Barbara S. Plake. "Cognitive Capacity Differences Among Writers." Journal of Educational Psychology 76 (1984): 820-34. (23, 48, 59).

Braddock, Richard. "Evaluation of Writing Tests." In Reviews of Selected Published Tests in English, Ed. Alfred H. Grommon. Urbana, Ill.: National Council of Teachers of English, 1976. (24, 49).

Breland, Hunter N., and Robert J. Jones. "Perceptions of Writing Skills." Written Communication 1.1 (Jan. 1984): 101-119. (ERIC No. EJ 293 034). (38, 53).

Britton, James, Tony Burgess, Nancy Martin, Alex McLeod, and Harold Rosen. The Development of Writing Abilities (11-18). London: Schools Council Publications, Macmillan Education LTD, 1975. Chapters 2, 8, and 9 abstracted. (6, 10, 14-15, 19-20, 47, 59).

-------. Language and Learning. London: Allen Lane The Penguin Press, 1970. Chapters 4 and 6 abstracted. (12, 18, 47, 58-59).

Burry, James, and Edys Quellmalz. "Assessing Students' Writing Skills: The CSE Expository and Narrative Rating Scales." Los Angeles: University of California, 1983. (ERIC Document Reproduction Service No. ED 238 942). (8, 10, 12, 50).

Camp, Gerald. "Study 3: Part II: A Field Test of Eight Prompts: Santa Clara County Writing Assessment." In Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment. Final Report. James Gray and Others. Berkeley, Calif.: University of California, 1982. (ERIC Document Reproduction Service No. ED 230 576). (36).

Caroll, Karen, and Sandra Murphy. "A Study of the Construction of the Meanings of a Writing Prompt by Its Authors, The Student Writers, and the Raters." In Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment. Final Report. James Gray and Others. Berkeley, Calif.: University of California, 1982. (ERIC Document Reproduction Service No. ED 230 576). (36, 53).

Chew, Charles R. "A Study to Determine the Relationship Between Indexes of Maturity and a Quality Grade in a Written Examination." Albany: New York State Education Dept., 1978. (ERIC Document Reproduction Service No. ED 178 914). (43, 54, 60).

--------. "A Writing Test: A Search for Competency: The New York Experience." Albany: New York State Education Dept., 1981. Paper read at International Conference on Language Problems and Public Policy, 16-19 Dec. 1981, Cancun, Mexico. (ERIC Document Reproduction Service No. ED 203 010). (25, 49).

Cooper, Charles R. "Holistic Evaluation of Writing." In

Evaluating Writing: Describing, Measuring, Judging. Ed. Charles R. Cooper and Lee Odell. Urbana, Ill.: National Council of Teachers of English, 1977. (6-10, 12-13).

-------. "Measuring Growth in Writing." English Journal 64.3 (Mar. 1975): 112-119. (11, 15-16).

-------, ed. The Nature and Measurement of Competency in English. Urbana, Ill.: National Council of Teachers of English, 1981. (40).

Diederich, Paul Bernard. Measuring Growth in English. Urbana, Ill.: National Council of Teachers of English, 1974. Chapter 3, 4, and Appendix A abstracted. (7-9, 11, 14, 29, 40-41, 50, 53-54, 59-60).

Dilworth, Jr. C.B., and R.W. Reising. "Validity in Composition Evaluation: Practicing What You Preach." Clearing House (Sept. 1979): 43-6. (8).

Fadiman, Clifton, and James Howard. Empty Pages: A Search for Writing Competence in School and Society. Belmont, Calif.: Fearon Pitman, 1979. (1, 11-12, 17, 47, 58).

Flinn, Jane Zeni. "An Accountability Model for School Writing Programs." English Education 16 (May, 1984): 101-10. (2).

Freedman, Sarah Warshauer, and Robert C. Calfee. "Holistic Assessment of Writing: Experimental Design and Cognitive Theory." In Research on Writing: Principles and Methods. Ed. Peter Mosenthal, Lynne Tamor, Sean A. Walmsley. New York: Longman, 1983. (1, 25).

-------. "Some Reasons for the Grades We Give Compositions." English Journal 71.7 (Nov. 1982): 86-89. ERIC No. EJ 269 776. (39, 53).

Gray, James, and Others. "Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment. Final Report." Berkeley, Calif.: University of California, 1982. (ERIC Document Reproduction Service No. ED 230 576). (13).

Gronlund, Norman E. Measurement and Evaluation in Teaching. 4th ed. New York: Macmillan Publishing Co., Inc., 1981. (8, 9-10, 13, 15, 41, 54, 60).

Hall, Janice K. Evaluating and Improving Written Expression: A Practical Guide for Teachers. Boston, Mass.: Allyn and Bacon, 1981. (9, 11, 23, 49).

Harris, Winifred Hall. "Teacher Response to Student Writing: A Study of the Response Patterns of High School English Teachers to Determine the Basis for Teacher Judgment of Student Writing." Research in the Teaching of English 11 (Fall, 1977): 175-85. (39, 53).

Hunt, Kellogg W. "Early Blooming and Late Blooming Syntactic Structures." In Evaluating Writing: Describing, Measuring, Judging. Ed. Charles R. Cooper and Lee Odell. Urbana, Ill.: National Council of Teachers of English, 1977. (15, 22, 48).

-------. Syntactic Maturity in School Children and Adults. Monographs of the Society for Research in Child Development, Serial No. 134, 35.1. Chicago: University

of Chicago Press, 1970. (30, 51).

Koffler, Stephen. "Assessing Students' Writing Skills: A Comparison of Direct and Indirect Methods." Trenton, New Jersey: New Jersey State Dept. of Education, 1984. Paper read at the Annual Conference of the American Educational Research Association, 1 Apr. 1984, New Orleans. (ERIC Document Reproduction Service No. ED 242 771). (46, 56, 60).

Langer, Judith A. "Effects of Topic Knowledge on the Quality and Coherence of Informational Writing." Berkeley: University of California, 1983. (ERIC Document Reproduction Service No. ED 234 418). (33, 52).

Lloyd-Jones, Richard. "Primary Trait Scoring." In Evaluating Writing: Describing Measuring Judging. Ed. Charles R. Cooper and Lee Odell. Urbana, Ill.: National Council of Teachers of English, 1977. (6-7, 28).

Macintosh, Helen, and Wilhelmina Hill. How Children Learn to Write. Bulletin No. 2--Place of--Subjects Series. Washington, D.C.: U.S. Dept. of Health, Education, and Welfare, Office of Education, 1953. (17, 47, 58).

Malgady, Robert G., and Peter R. Barcher. "Some Information Processing Models of Creative Writing." Journal of Educational Psychology 71.5: 717-25. (30, 51).

Myers, Miles. "A Procedure for Writing Assessment and

Holistic Scoring." Oakland, Calif: Bay Area Writing Project, 1980. (ERIC Document Reproduction Service No. ED 193 676). Chapters 1 and 3 abstracted. (6-7, 13-14, 32, 37, 42, 52-54, 59).

Nyberg, Verner R. and Adell N. Nyberg. "Alberta Essay Scales: Models." Edmonton, Canada: University of Alberta, 1982. (ERIC Document Reproduction Service No. ED 220 483). (42, 54).

-------. Technical Report on Alberta Essay Scales: Models." Edmonton, Canada: University of Alberta, 1982. (ERIC Document Reproduction Service No. ED 220 484). (27, 50).

Odell, Lee. "Defining and Assessing Competence in Writing." In The Nature and Measurement of Competency in English. Ed. Charles R. Cooper. Urbana, Ill.: National Council of Teachers of English, 1981. (37, 53).

-------. "Measuring Changes in Intellectual Processes as One Dimension of Growth in Writing." In Evaluating Writing: Describing, Measuring, Judging. Ed. Charles R. Cooper and Lee Odell. Urbana, Ill.: National Council of Teachers of English, 1977. (31, 51).

-------. "Redefining Mature Writing." Albany: State Univ. of New York, 1979. Paper read at the Annual Meeting of the Canadian Council of Teachers of English, 8-11 May, 1979, Ottawa Canada. (ERIC Document Reproduction Service No. ED 173 781). (22, 48).

O'Neal, Marcia R., James E. McLean, and George E. McCormick.

"The Use of the Syntactic Density Score as an Evaluative Criterion Measure." Tuscaloosa: University of Alabama, 1983. Paper read at the Annual Meeting of the Mid-South Educational Research Association, 16-18 Nov. 1983, Nashville, Tenn. (ERIC Document Reproduction Service No. ED 237 558). (31, 51).

Plasse, Lorraine A. "The Influence of Audience on the Assessment of Student Writing." Springfield, Mass.: Springfield Public Schools, 1982. (ERIC Document Reproduction Service No. ED 229 760). (35, 52, 59).

Prater, Doris. "The Effects of Writing Task on Mastery Classification of High School Students." Clear Lake City, Texas: University of Houston, 1982. Paper read at the Annual Meeting of the Southwest Educational Research Association, Feb. 1982, Austin, Texas. (ERIC Document Reproduction Service No. ED 215 367). (35, 52).

Quellmalz, Edys. "Designing Writing Assessments: Balancing Fairness, Utility, and Cost." Los Angeles: University of California, 1982. (ERIC Document Reproduction Service No. ED 228 270). (2).

-------, and Frank Cappell. "Defining Writing Domains: Effects of Discourse and Response Mode." Los Angeles: University of California, 1979. (ERIC Document Reproduction Service No. ED 212 661). (34, 52, 59).

Rabianski, Nancyanne. "Scoring Writing With an Informative Aim." Document reviewed in "Measures for Research

and Evaluation in the English Language Arts," Vol. 2, The Research Instruments Project, Urbana, Ill.: Committee on Research, National Council of Teachers of English, 1977. (ERIC Document Reproduction Service No. ED 236 638). (28, 50).

Ruth, Leo. "Study 1: Designing Prompts for Holistic Writing Assessments: Knowledge From Theory, Research and Practice." In Properties of Writing Tasks: A Study of Alternative Procedures for Holistic Writing Assessment. Final Report. Berkeley, Calif.: University of California, 1982. (ERIC Document Reproduction Service No. ED 230 576). (32, 52, 58).

Sloan, Charles A. and Iris McGinnis. "The Effect of Handwriting on Teachers' Grading of High School Essays." DeKalb, Ill.: Northern Illinois University, 1978. (ERIC Document Reproduction Service No. ED 220 836). (40, 53).

Smith, Laura Spooner. "Investigation of Writing Assessment Strategies. Studies in Measurement and Methodology. Work Unit 1: Design and Use of Tests." Los Angeles: University of California, 1978. (ERIC Document Reproduction Service No. ED 213 727). (45, 55, 58, 60-61).

Spencer, Ernest. "Folio Assessment or External Examinations? An Investigation into Alternative Means of Assessing SCE Ordinary Grade English." Report to the Scottish Certificate of Education Examination Board from the Scottish Council for Research in Education,

1979. (ERIC Document Reproduction Service No. ED 194 555). (26).

Stotsky, Sandra. "Examining Word Use to Assess Growth in Essay Writing." This essay is the second revision of a paper presented at the 6th Annual Conference on Language Development, Oct., 1981, Boston. (ERIC Document Reproduction Service No. ED 211 980). (44, 55, 60).

Tamor, Lynne and James T. Bond. "Text Analysis: Inferring Process from Product." In Research on Writing: Principles and Methods. Ed. Peter Mosenthal, Lynne Tamor, and Sean Walmsley. New York: Longman, 1983. (2).

Thrall, William Flint, Addison Hibbard, and C. Hugh Holman. A Handbook to Literature. New York: The Odyssey Press, 1960. (10-11).

Tierney, Robert J. and James H. Mosethal. "Cohesion and Textual Coherence." Research in the Teaching of English 17.3 (Oct. 1983): 215-29. (ERIC No. EJ 289 567). (43, 55, 60).

Veal, Ramon L., and Sally Ann Hudson. "Direct and Indirect Measures for Large-Scale Evaluation of Writing." Research in the Teaching of English 17 (Oct. 1983): 290-96. (ERIC No. EJ 289 573). (45, 55, 58, 61).

Wade, Barrie. "Responses to Written Work: The Possibilities of Utilizing Pupils Perceptions." Educational Review 30.2 (Jun. 1978): 149-58. (ERIC No. EJ 188 175). (41, 54, 60).

Webster's New Twentieth Century Dictionary, Unabridged. 2nd ed. USA: William Collins Publishers, 1980. (7, 9, 11).

White, E.M. "Holisticism." College Composition and Communication 35 (Dec. 1984): 400-9. (24).

Winters, Lynn. "The Effects of Differing Response Criteria on the Assessment of Writing Competence." Los Angeles: University of California, 1978. (ERIC Document Reproduction Service No. ED 212 659). (13, 44, 55, 60-61).